Once-militant sea defenders offer their fleet to science p. 18

Bioresorbable coolers to block nerve pain pp. 28 & 109

Oceania was peopled in multiple waves by male migrants p. 72

Science

$15
1 JULY 2022
science.org
AAAS



UNDERSEA EXPLOSION

The volcanic eruption in Tonga heard around the world pp. 30, 91, & 95

VISIT…

LANZAROTE Caliente.com

EDITORIAL

Science, health, and truth

The past 2 years have been a strange and difficult time for the world. The COVID-19 pandemic has brought illness, hospitalization, and death near to many people. In the United States, people are divided not only on what they should do but also on what constitutes the facts. Many are seemingly in an alternative world, driven by disinformation, conspiracy theories, and anti-science beliefs. How can health and medical leaders do their jobs while trying to cope with a polarized public? They must be more effective in explaining and persuading the public on matters of science and health. This will require better clarification of two things to the public—the roles of science and politics in public policy decisions, and the means by which scientific truth is established and updated.

There is a continuing strand of argument from many in the scientific community that "we ought to get politics out of public health." And many politicians have said that scientists are not the ones to make public policy decisions related to public health or other issues. The reality is that both science and politics are essential for public health to work well. Scientists inform public understanding of the patterns of health and illness in populations, especially when epidemics and pandemics strike. And politics—the way decisions are made in a democratic society—is vital for acting on the information and insights that the scientific community provides for the benefit of everyone.

> **"How can health and medical leaders do their jobs while trying to cope with a polarized public?"**

In addition to clarifying the separate roles of science and politics, society needs to understand better how scientific truth is established and updated. It is based on verified and reproducible facts. The scientific method of gathering data, debating various formulations of the information, and arriving at consensus understandings of what is "true" about a particular matter has been the bedrock for establishing scientific truth for centuries. But in the past few years, political leaders, media personalities, and ordinary citizens have proclaimed their own "alternative facts," disparaging the idea that anyone should question their "facts" and theories.

In his recent book, *The Constitution of Knowledge: A Defense of Truth*, Jonathan Rauch clearly and convincingly describes and defends the Enlightenment-born process of weighing evidence and committing to truth. He explains that the endeavor is founded on two arguments: The truth is based on what is known so far and the truth applies to everyone. Rauch argues that this has produced a process for certifying knowledge—a public system for adjudicating differences of belief and perception and for developing shared and warranted conclusions about truth. This system provides three public goods: knowledge (distinguishing reality from nonreality), freedom (encouraging rather than repressing human autonomy, creativity, and empowerment), and peace (rewarding social conciliation, maximizing the number of disagreements that are resolvable, and compartmentalizing and marginalizing disagreements that cannot be resolved).

Given that truth is based on what is known so far, it is always provisional. We likely will know more tomorrow than today, and we will then have to revise our conclusions. That is just the scientific method at work. Though science and politics have separate roles, once this system "gets the facts straight," scientists and political leaders can work constructively together for the common good.

This knowledge-certifying system is under concerted attack today, most notably in polarized political conflicts, including about masks and vaccines, climate change, and gun violence. Restoring confidence in messages regarding science for the public good will be challenging, but it can only be done if there is an effort to explain, defend, and reinforce this public system for shepherding new knowledge. This needs to include widespread discussions across the scientific community and in the public at large, leading to a broad recommitment to the scientific method.

The challenges of communicating messages about science, health, and truth are among the often-mentioned reasons for the fatigue and discouragement that public health and medical leaders face today. And that same frustration is transferred to those who seek to engage the wider public with these messages.

Overcoming these challenges will surely be difficult, but with concerted effort it can be done, as it has before—for example, in efforts to promote widespread childhood immunizations. And this requires a never-ending commitment, as we are learning every day.

–William L. Roper



William L. Roper is a professor of medicine and public health at the University of North Carolina–Chapel Hill, Chapel Hill, NC, USA. roper@unc.edu

PHOTO: UNC HEALTH

10.1126/science.add6477

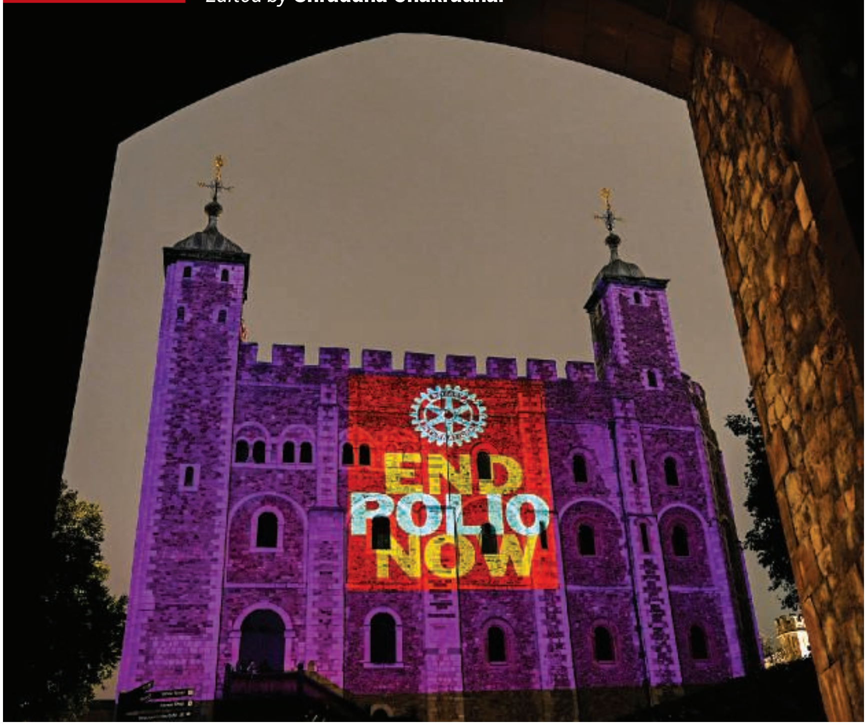
NEWS

> "The already broken healthcare system is going to collapse under this decision. … Pregnancy is too dangerous."

Tweet by Jocelyn J. Fitzgerald, a urogynecologist at the University of Pittsburgh, on the U.S. Supreme Court's reversal last week of its *Roe* v. *Wade* abortion decision.

IN BRIEF

Edited by **Shraddha Chakradhar**



The Tower of London was lit up to mark World Polio Day in October 2021.

PHOTO: AARON CHOWN/PA IMAGES/GETTY IMAGES

PUBLIC HEALTH

Poliovirus in U.K. sewage raises alarm

Officials in the United Kingdom declared a "national incident" last week after poliovirus was detected in samples taken between February and June from one of London's main sewage treatment plants. The source is still a mystery, but it is likely that someone from outside the United Kingdom who recently received the oral polio vaccine—which uses a live, but weakened, poliovirus—is shedding the pathogen in their feces. No polio cases have been found and such sewage detections are not unusual—signs of the virus usually disappear quickly. This time, however, it continued to appear over many months, and several closely related versions of the virus were in recent samples. The genetic changes suggest the virus is evolving, a sign it may be spreading in a small number of people. The United Kingdom's generally high polio vaccine coverage should limit further transmission, infectious disease experts say—but they worry lower rates in some London communities might leave children there vulnerable.

Launch of NASA's Psyche delayed

PLANETARY SCIENCE | Psyche, the $967 million NASA mission to explore the strange metal asteroid of the same name, will not launch this fall because of delays in testing its navigation software, the agency announced last week. The mission, developed by the Jet Propulsion Laboratory, is likely to launch next year, but could also slip to 2024; either date would see its arrival at the main-belt asteroid, which was set for 2026, fall back by 3 years or more. An independent panel will review what went wrong and decide whether the mission will continue—although cancellation is unlikely. The delay should not limit the science conducted by the spacecraft, which aims to investigate whether the asteroid is the leftover core of an early planetary embryo.

NCI to study cancer blood tests

MEDICINE | The National Cancer Institute (NCI) plans to evaluate the growing number of tests from companies that screen for multiple cancers by detecting traces of tumor DNA and proteins in a person's blood. Last month, NCI advisers endorsed a $75 million, 4-year pilot study that will enroll 24,000 seemingly cancer-free volunteers starting in 2023 or 2024. Some would only get standard cancer screening, such as mammograms, whereas others would also receive one of the experimental blood tests. If the pilot study goes well, NCI expects to launch a bigger trial, with perhaps 300,000 participants ages 45 to 70, that would last 7 to 8 years. Such a long-term study could reveal whether these tests, which can return false positives and lead to unnecessary procedures on tiny, slow-growing tumors, ultimately save lives.

Austria rescinds vaccine mandate

COVID-19 | Austria, one of the first countries in Europe to pass a COVID-19 vaccine mandate, officially rescinded the requirement last week. The law, passed in November 2021 with support across the political spectrum, required all adults to be vaccinated by February or face fines of up to €3600 ($3800). However, the government never enforced it. By March,



The Citadel ruins in Bears Ears National Monument will get tribal oversight.

CONSERVATION

Native tribes and United States enter park management deal

The Bureau of Land Management (BLM) and the U.S. Forest Service signed a first-of-its-kind agreement last week to comanage Bears Ears National Monument in southeastern Utah with five Native American tribes. The five tribes have long wanted to help oversee this culturally significant monument, a 550,000-hectare expanse pocked by burnt-sienna buttes, 1000-year-old Indigenous dwellings, and ceremonial sites. The comanagement agreement is meant to incorporate the traditional knowledge and expertise of the tribes—the Hopi, Navajo Nation, Ute Mountain Ute Tribe, Ute Indian Tribe of the Uintah and Ouray Reservation, and the Pueblo of Zuni—and center their voices in land use decisions. The Indigenous tribes will now submit a draft land management plan to BLM in the coming weeks, which could then take up to 18 months to finalize.

the government said the Omicron variant, which evades vaccine protection better than previous versions of SARS-CoV-2, meant the mandate was no longer justified. Last week, Austria's health minister said the law had failed to persuade the vaccine-hesitant and had been more divisive than helpful. He said the government would now focus on explaining why receiving a booster dose is important. Austria has one of the lower vaccination rates in Europe, with fewer than 75% of the population fully vaccinated.

FDA tries to stop Juul sales

HEALTH POLICY | The e-cigarette manufacturer Juul last week won a temporary reprieve from a U.S. Food and Drug Administration (FDA) order banning the company from selling its products in the country. The 23 June FDA order would affect all Juul products in the United States, including flavored cartridges and pods for vaping. It came as part of a larger agency effort to curb high rates of teen vaping and other public health crises fueled by nicotine use. Earlier last week, for instance, FDA announced it would be mandating lower nicotine levels in traditional cigarettes. A federal appeals court granted Juul time to file a formal challenge to the FDA order, which the company asserts is the product of political maneuvering.

BY THE NUMBERS

U.S. science spending bills advance

The U.S. House of Representatives last week released proposed 2023 budgets for key research agencies (see below). The Senate has yet to release its proposals, and final numbers likely won't be set until late this year.

	2022 ($ BILLIONS)	BIDEN 2023 REQUEST	HOUSE BILLS
National Institutes of Health	45	50.5	47.5
National Science Foundation	8.8	10.5	9.6
NASA science	7.6	8	7.9
Department of Energy science	7.5	7.8	8
DARPA	3.8	4.1	4
U.S. Geological Survey	1.4	1.7	1.6
NIST core labs	0.85	0.98	0.95
EPA science	0.73	0.86	0.87

The Lancet tops medical citations

PUBLISHING | The title of most cited medical journal has changed hands for the first time in 45 years. In 2021, *The Lancet* overtook *The New England Journal of Medicine* (*NEJM*) as the publication with the highest impact factor, according to a report released this week by the analytics firm Clarivate. The company manages the Web of Science database used to generate this controversial citation measure, and *NEJM* has dominated the general and internal medicine category since Clarivate's Journal Citation Reports launched in the 1970s. High-profile pandemic research may explain the shift: *The Lancet* published three of the 10 most cited articles in 2021, all of them related to COVID-19.

CREDITS: (PHOTO) BOB WICK/BUREAU OF LAND MANAGEMENT/CC BY 2.0; (DATA) HOUSE OF REPRESENTATIVES APPROPRIATIONS COMMITTEE

IN DEPTH



EUROPE

United Kingdom set to abandon EU funding and go it alone

Horizon Europe grants held hostage over Brexit dispute

By **Cathleen O'Grady**

A few months ago, Teresa Thurston, a cellular microbiologist at Imperial College London, could not have imagined losing her €1.5 million European research grant. But the United Kingdom's role in the European Union's €95 billion Horizon Europe funding program is now crumbling thanks to lingering Brexit disputes, forcing many U.K. grant winners like Thurston to give up grants they thought they could count on.

This week, she and 142 other U.K. grant winners faced a deadline from the European Research Council (ERC), which disburses some of the Horizon Europe money. By 29 June, they had to choose between moving to a European host institution to take up their funding or remaining in the United Kingdom and giving up the prestigious ERC grant. One day before the deadline, 18 U.K.-based grant winners had told ERC they planned to leave for Europe, whereas 20, including Thurston, rejected their ERC grants to stay in the United Kingdom. The remainder had not yet told ERC of their decision.

The impasse over U.K. participation in Horizon Europe deepened last month after the U.K. Parliament began pushing through legislation that would unilaterally override part of the 2020 Brexit agreement that deals with trade in Northern Ireland. This prompted legal action from the European Union, which had already been holding cross-border scientific funding hostage over the matter, and signaled that a resolution was far from near. The U.K. science minister on 15 June said he would unveil details of a "Plan B" Horizon alternative in July, adding that the program will likely begin in September.

In the meantime, a transitional program promises to make up for Horizon grants already won by U.K. researchers. Thurston plans to accept the government's offer, but she says this backstop doesn't carry the same prestige or networking opportunities. "It's not the same," she says.

ERC President Maria Leptin says she "profoundly" regrets the situation. "I sincerely hope that our political leaders will find a path to allow us to continue working in the future with the research community in the U.K."

Greg Clark, a U.K. lawmaker and member of a parliamentary science committee, says it's wrong for the European Union to "weaponize science" to gain advantage in other political arguments. Other components of the Brexit agreement, such as the cross-border transfer of personal data, have gone ahead despite the Northern Ireland dispute. And although damage to U.K. science is considerable, so is damage to science within the EU, he says: "The longer term consequences of excluding excellent U.K. institutions and researchers from collaboration with European institutions can only be to the detriment of both sides." A European Commission spokesperson said in an email to *Science* that, given the turmoil, the original Brexit agreement does not oblige the EU to allow U.K. participation in Horizon Europe, nor does it impose a deadline.

The Brexit deal was supposed to preserve U.K. participation. In an arrangement similar to those for other non-EU countries that receive Horizon Europe funds, such as Israel, Turkey, and Norway, the United Kingdom would pay approximately £15 billion into the program over its 7-year period, allowing its researchers to compete for funding. The European Parliament still had to ratify the agreement, but U.K. researchers began to apply for—and win—grants in the first funding rounds of the program.

In April, however, the political impasse over Northern Ireland was ongoing and the Horizon deal was still not ratified. Facing the need to disburse its first round of grants, ERC emailed grant winners asking them to submit letters from Horizon-eligible host institutions within 2 months—or forfeit their grants.

It's "massively destabilizing," says Peter Mason, head of global research and innovation policy at Universities UK. "There are these great collaborations—scientific networks—that have been established for decades in some cases, which are being undermined." James Wilsdon, a research policy expert at the University of Sheffield, adds that "all of this uncertainty is just like acid on collaboration."

For Thurston, moving to Europe was not practical, because her collaborators were all

ILLUSTRATION: DAVIDE BONAZZI

Corrected 30 June 2022. See full text.

in the United Kingdom. She also did not want to uproot her family, and she points out that a short-notice move to Europe would be hard for researchers taking care of children, which means women likely face greater impact.

Thiemo Fetzer, an economist at the University of Warwick, reluctantly made the decision to move. A collaborator on his project is employed by the European Commission and would not be able to sign a contract with a U.K.-based institution. Fetzer will now divide his time between the United Kingdom and his European host, which he declined to name because he was still finalizing the agreement. "It's devastating," Fetzer says. Winning an ERC grant is "supposed to be a high point of people's academic careers, and it's made it really, really difficult."

European researchers aren't happy either, says Thomas Jørgensen, director of policy coordination and foresight at the European University Association. The European Union's lack of transparency and poor communication has damaged the profile of Horizon Europe, he says. European researchers were encouraged to join U.K.-led consortia, and then faced chaos as the U.K. partners were no longer able to take their intended leading roles. Some top researchers now feel that EU funding is too bureaucratic and difficult to be worth bothering with again, Jørgensen says.

In theory, the United Kingdom could join Horizon Europe at any time in the coming 7 years if political disputes over Brexit were settled. And Horizon association remains "Plan A," the U.K. science minister, George Freeman, said in a 15 June parliamentary committee hearing. He could not be reached for comment, but a department spokesperson said, "The U.K. is pushing for association to Horizon Europe as swiftly as possible."

But Freeman told the committee that if "the phone does not ring" soon, then "Plan B" will begin in September, with the government diverting the £15 billion intended for Horizon into a U.K.-based international program. Freeman described plans to replicate the structure of the Horizon funding scheme, with grants not only for basic research, similar to ERC's, but also funding for risky, applied, and industrial research.

Wilsdon adds that the bulk of Horizon Europe funding goes toward international research consortia that often involve industrial collaborators. U.K.-based researchers have already been forced to step down from their coordinating roles in Horizon-funded consortia, and Wilsdon thinks it would be tough to re-create these in Freeman's "Plan B."

Beth Thompson, a policy director at the Wellcome Trust, shares the concern. "All those researchers' brains coming together to do something brilliant—that can't be reproduced at a domestic level," she says. ■

SCIENTIFIC PUBLISHING

Silence greets requests to flag retracted studies

Authors and editors ignored warnings about citing noted fraudster, exposing a problem in scholarly publishing

By **Jeffrey Brainard**

Alison Avenell spent years collecting evidence that Yoshihiro Sato, a now-deceased nutritional researcher in Japan, was among the most prolific fraudsters known to science. After journals investigated the findings by Avenell, a clinical nutritionist at the University of Aberdeen, and her colleagues, they retracted more than two dozen papers Sato had co-authored. Many had reported findings from clinical trials that could have led physicians to incorrectly treat patients suffering from osteoporosis and other disorders.

But the retractions, which began in 2015, didn't mean the papers were gone for good, or that their influence waned.

Avenell noticed many journal articles that cited one or more of Sato's 27 retracted papers did not warn readers that they referenced tainted work. Worse, she and colleagues reported recently, the flawed findings were cited in 88 articles, published from 2003 to 2020, that are systematic reviews and clinical guidelines—potentially influential publications that often help guide medical treatments. Avenell wondered: Would the authors and editors of these papers take action if alerted to the retractions of Sato's work?

For the most part, she found, the answer was no.

Her team contacted the authors of 86 of the citing papers—and sometimes the editors, too. After a year, however, journals had posted notices or letters for just eight of those papers informing readers that they cited retracted work, the researchers reported in late May in *Accountability in Research*. In five of those cases the announcement wasn't linked to the paper, leaving readers in the dark. (A ninth review was itself retracted.)

The saga provides an unusually methodical case study of what some call "zombie papers." Even after they are retracted—publishing's death sentence—these papers live on thanks to citations. And that could have real-world consequences, the study suggests. It found 39 of the 88 citing papers had drawn conclusions that, if Sato's retracted papers were left out of the analysis, were likely to be substantially weaker. Journals flagged just four of these weakened studies for citing retracted papers.

The study's findings are "unfortunately very consistent" with others going back to the 1990s, says Ivan Oransky, co-editor of Retraction Watch, which reports on retracted papers and tracks them in a public database. A 1998 investigation in *JAMA*, for example, found that 94% of 299 citations to retracted articles still listed in the MEDLINE database did not note the work had been retracted. And "most editors do not seem to make correcting the record a priority," Oransky says.

Avenell took a very rigorous approach to documenting the problem, Oransky says. For example, her team emailed queries to authors and journals using a randomized, controlled trial design. For some papers, the researchers contacted only the corresponding authors of the evidence syntheses. For others, they contacted an additional two co-authors and sometimes also the journal's editor-in-chief.

For half of the 86 papers, they got no response. (Looping in the editor didn't increase the response rate.) Some authors who did respond said they didn't plan to amend their papers because, for example, the publication was too old,

88

Medical evidence syntheses citing one or more of 27 retracted clinical trial papers co-authored by scientific fraudster Yoshihiro Sato

39

Syntheses out of the 88 with findings that, if the retracted paper had been excluded, would likely be weaker

4

Syntheses out of the 39 whose journals added a notice saying it cited a retracted paper, after the authors were notified about the retraction

DATA: A. AVENELL, M. BOLLAND, G. GAMBLE, AND A. GREY, *ACCOUNTABILITY IN RESEARCH*, DOI: 10.1080/08989621.2022.2082290 (2022)

Corrected 30 June 2022. See full text.

in the United Kingdom. She also did not want to uproot her family, and she points out that a short-notice move to Europe would be hard for researchers taking care of children, which means women likely face greater impact.

Thiemo Fetzer, an economist at the University of Warwick, reluctantly made the decision to move. A collaborator on his project is employed by the European Commission and would not be able to sign a contract with a U.K.-based institution. Fetzer will now divide his time between the United Kingdom and his European host, which he declined to name because he was still finalizing the agreement. "It's devastating," Fetzer says. Winning an ERC grant is "supposed to be a high point of people's academic careers, and it's made it really, really difficult."

European researchers aren't happy either, says Thomas Jørgensen, president of the European University Association. The European Union's lack of transparency and poor communication has damaged the profile of Horizon Europe, he says. European researchers were encouraged to join U.K.-led consortia, and then faced chaos as the U.K. partners were no longer able to take their intended leading roles. Some top researchers now feel that EU funding is too bureaucratic and difficult to be worth bothering with again, Jørgensen says.

In theory, the United Kingdom could join Horizon Europe at any time in the coming 7 years if political disputes over Brexit were settled. And Horizon association remains "Plan A," the U.K. science minister, George Freeman, said in a 15 June parliamentary committee hearing. He could not be reached for comment, but a department spokesperson said, "The U.K. is pushing for association to Horizon Europe as swiftly as possible."

But Freeman told the committee that if "the phone does not ring" soon, then "Plan B" will begin in September, with the government diverting the £15 billion intended for Horizon into a U.K.-based international program. Freeman described plans to replicate the structure of the Horizon funding scheme, with grants not only for basic research, similar to ERC's, but also funding for risky, applied, and industrial research.

Wilsdon adds that the bulk of Horizon Europe funding goes toward international research consortia that often involve industrial collaborators. U.K.-based researchers have already been forced to step down from their coordinating roles in Horizon-funded consortia, and Wilsdon thinks it would be tough to re-create these in Freeman's "Plan B."

Beth Thompson, a policy director at the Wellcome Trust, shares the concern. "All those researchers' brains coming together to do something brilliant—that can't be reproduced at a domestic level," she says. ■

SCIENTIFIC PUBLISHING

Silence greets requests to flag retracted studies

Authors and editors ignored warnings about citing noted fraudster, exposing a problem in scholarly publishing

By **Jeffrey Brainard**

Alison Avenell spent years collecting evidence that Yoshihiro Sato, a now-deceased nutritional researcher in Japan, was among the most prolific fraudsters known to science. After journals investigated the findings by Avenell, a clinical nutritionist at the University of Aberdeen, and her colleagues, they retracted more than two dozen papers Sato had co-authored. Many had reported findings from clinical trials that could have led physicians to incorrectly treat patients suffering from osteoporosis and other disorders.

But the retractions, which began in 2015, didn't mean the papers were gone for good, or that their influence waned.

Avenell noticed many journal articles that cited one or more of Sato's 27 retracted papers did not warn readers that they referenced tainted work. Worse, she and colleagues reported recently, the flawed findings were cited in 88 articles, published from 2003 to 2020, that are systematic reviews and clinical guidelines—potentially influential publications that often help guide medical treatments. Avenell wondered: Would the authors and editors of these papers take action if alerted to the retractions of Sato's work?

For the most part, she found, the answer was no.

Her team contacted the authors of 86 of the citing papers—and sometimes the editors, too. After a year, however, journals had posted notices or letters for just eight of those papers informing readers that they cited retracted work, the researchers reported in late May in *Accountability in Research*. In five of those cases the announcement wasn't linked to the paper, leaving readers in the dark. (A ninth review was itself retracted.)

The saga provides an unusually methodical case study of what some call "zombie papers." Even after they are retracted—publishing's death sentence—these papers live on thanks to citations. And that could have real-world consequences, the study suggests. It found 39 of the 88 citing papers had drawn conclusions that, if Sato's retracted papers were left out of the analysis, were likely to be substantially weaker. Journals flagged just four of these weakened studies for citing retracted papers.

The study's findings are "unfortunately very consistent" with others going back to the 1990s, says Ivan Oransky, co-editor of Retraction Watch, which reports on retracted papers and tracks them in a public database. A 1998 investigation in *JAMA*, for example, found that 94% of 299 citations to retracted articles still listed in the MEDLINE database did not note the work had been retracted. And "most editors do not seem to make correcting the record a priority," Oransky says.

Avenell took a very rigorous approach to documenting the problem, Oransky says. For example, her team emailed queries to authors and journals using a randomized, controlled trial design. For some papers, the researchers contacted only the corresponding authors of the evidence syntheses. For others, they contacted an additional two co-authors and sometimes also the journal's editor-in-chief.

For half of the 86 papers, they got no response. (Looping in the editor didn't increase the response rate.) Some authors who did respond said they didn't plan to amend their papers because, for example, the publication was too old,

88
Medical evidence syntheses citing one or more of 27 retracted clinical trial papers co-authored by scientific fraudster Yoshihiro Sato

39
Syntheses out of the 88 with findings that, if the retracted paper had been excluded, would likely be weaker

4
Syntheses out of the 39 whose journals added a notice saying it cited a retracted paper, after the authors were notified about the retraction

DATA: A. AVENELL, M. BOLLAND, G. GAMBLE, AND A. GREY, *ACCOUNTABILITY IN RESEARCH*, DOI: 10.1080/08989621.2022.2082290 (2022)

or they didn't have time to do a reanalysis. Some asserted that the elimination of a single, retracted study likely would not have changed their overall findings. There is some evidence for that position. A 2021 study in *Accountability in Research* led by Daniele Fanelli of the London School of Economics and Political Science examined 50 meta-analyses of clinical treatments. The conclusions of those that cited retracted work and those that didn't were statistically similar.

The studies examined by Avenell's team that were weakened by retracted findings could have put patients at risk. One of those reviews, showing that vitamin K helps prevent bone fractures, was the basis of 2011 and 2015 guidelines issued in Japan that recommend the supplement for people at risk. Omitting Sato's studies made the reported benefit statistically nonsignificant. The guidelines' sponsor, the Japan Osteoporosis Foundation, was among those that did not respond to the team's queries.

Even if the inclusion of a retracted study doesn't change the bottom line, Avenell argues, journals and authors have an obligation to say so publicly. "You need to reassure your readers" about a paper's validity, she says.

Avenell is scheduled to discuss the study—co-authored by Mark Bolland, Greg Gamble, and Andrew Grey of the University of Auckland—in September at the International Congress on Peer Review and Scientific Publication. The work, she says, was spurred by "my frustration with the slow process of correcting the literature that has integrity issues and to demonstrate the potential adverse consequences if it is not corrected promptly."

There are signs the research community is beginning to take such concerns more seriously. Several bibliographic databases—including EndNote, LibKey, Papers, and Zotero—now note papers that are included in Retraction Watch's database of retractions, which debuted publicly in 2018. (The popular Google Scholar search engine does not flag retractions.) The International Committee of Medical Journal Editors recommends journal editors routinely check to see whether submitted manuscripts cite retracted papers. And in 2021, Cochrane, a nonprofit international network that promotes evidence-based medicine, began to attach a warning to any of its published systematic reviews that cite retracted studies. Cochrane asks the authors of the flagged reviews to reconsider their work; then the organization decides whether to withdraw the analysis or publish an updated version. ■



The Vestmanna cliffs of the Faroe Islands formed when Greenland tore away from Europe 56 million years ago.

CLIMATE CHANGE

Hidden carbon layer sparked ancient bout of global warming

Deep carbon exhumed by volcanic rift between Greenland and Europe implicated in 56-million-year-old hothouse

By **Paul Voosen**

There is no perfect parallel in Earth's past for present-day climate change—human-driven warming is simply happening too fast and furiously. The closest analog came 56 million years ago, when over the course of 3000 to 5000 years, greenhouse gases soared in the atmosphere, causing at least 5°C of warming and pushing tropical species to the poles.

The cause of the Paleocene-Eocene Thermal Maximum (PETM) has long been debated, with researchers invoking exotic mechanisms such as catastrophic releases of methane from the sea floor or even asteroid strikes. But over the past few years, evidence has mounted for a more prosaic culprit: carbon-spewing volcanoes that emerged underneath Greenland as it tore away from Europe. Now, researchers have found signs of an effect that would have supercharged the warming effect of the volcanoes, making them a stronger suspect. The underside of Greenland is thought to be encrusted with carbon-rich rocks, like barnacles on the keel of a ship. When it tore away from Europe in the rifting, the rocks might have liberated a gusher of carbon dioxide (CO_2), says Thomas Gernon, a geologist at the University of Southampton and leader of the new study. "It's a perfect storm of conditions."

The PETM has long fascinated paleoclimatologists. "Since dinosaurs kicked the bucket, this is the biggest global warming event we have," says Pincelli Hull, a paleoclimate scientist at Yale University. It can yield clues to how quickly Earth warms as greenhouse gas levels rise and how climate extremes alter ecosystems. But the comparison to today isn't exact. Although the total release of carbon during the PETM

PHOTO: STEVE_IS_ON_HOLIDAY/ISTOCK.COM

or they didn't have time to do a reanalysis. Some asserted that the elimination of a single, retracted study likely would not have changed their overall findings. There is some evidence for that position. A 2021 study in *Accountability in Research* led by Daniele Fanelli of the London School of Economics and Political Science examined 50 meta-analyses of clinical treatments. The conclusions of those that cited retracted work and those that didn't were statistically similar.

The studies examined by Avenell's team that were weakened by retracted findings could have put patients at risk. One of those reviews, showing that vitamin K helps prevent bone fractures, was the basis of 2011 and 2015 guidelines issued in Japan that recommend the supplement for people at risk. Omitting Sato's studies made the reported benefit statistically nonsignificant. The guidelines' sponsor, the Japan Osteoporosis Foundation, was among those that did not respond to the team's queries.

Even if the inclusion of a retracted study doesn't change the bottom line, Avenell argues, journals and authors have an obligation to say so publicly. "You need to reassure your readers" about a paper's validity, she says.

Avenell is scheduled to discuss the study—co-authored by Mark Bolland, Greg Gamble, and Andrew Grey of the University of Auckland—in September at the International Congress on Peer Review and Scientific Publication. The work, she says, was spurred by "my frustration with the slow process of correcting the literature that has integrity issues and to demonstrate the potential adverse consequences if it is not corrected promptly."

There are signs the research community is beginning to take such concerns more seriously. Several bibliographic databases—including EndNote, LibKey, Papers, and Zotero—now note papers that are included in Retraction Watch's database of retractions, which debuted publicly in 2018. (The popular Google Scholar search engine does not flag retractions.) The International Committee of Medical Journal Editors recommends journal editors routinely check to see whether submitted manuscripts cite retracted papers. And in 2021, Cochrane, a nonprofit international network that promotes evidence-based medicine, began to attach a warning to any of its published systematic reviews that cite retracted studies. Cochrane asks the authors of the flagged reviews to reconsider their work; then the organization decides whether to withdraw the analysis or publish an updated version. ■



The Vestmanna cliffs of the Faroe Islands formed when Greenland tore away from Europe 56 million years ago.

CLIMATE CHANGE

Hidden carbon layer sparked ancient bout of global warming

Deep carbon exhumed by volcanic rift between Greenland and Europe implicated in 56-million-year-old hothouse

By **Paul Voosen**

There is no perfect parallel in Earth's past for present-day climate change—human-driven warming is simply happening too fast and furiously. The closest analog came 56 million years ago, when over the course of 3000 to 5000 years, greenhouse gases soared in the atmosphere, causing at least 5°C of warming and pushing tropical species to the poles.

The cause of the Paleocene-Eocene Thermal Maximum (PETM) has long been debated, with researchers invoking exotic mechanisms such as catastrophic releases of methane from the sea floor or even asteroid strikes. But over the past few years, evidence has mounted for a more prosaic culprit: carbon-spewing volcanoes that emerged underneath Greenland as it tore away from Europe. Now, researchers have found signs of an effect that would have supercharged the warming effect of the volcanoes, making them a stronger suspect. The underside of Greenland is thought to be encrusted with carbon-rich rocks, like barnacles on the keel of a ship. When it tore away from Europe in the rifting, the rocks might have liberated a gusher of carbon dioxide (CO_2), says Thomas Gernon, a geologist at the University of Southampton and leader of the new study. "It's a perfect storm of conditions."

The PETM has long fascinated paleoclimatologists. "Since dinosaurs kicked the bucket, this is the biggest global warming event we have," says Pincelli Hull, a paleoclimate scientist at Yale University. It can yield clues to how quickly Earth warms as greenhouse gas levels rise and how climate extremes alter ecosystems. But the comparison to today isn't exact. Although the total release of carbon during the PETM

PHOTO: STEVE_IS_ON_HOLIDAY/ISTOCK.COM

exceeded the total of today's known oil and gas reserves, it was slower than today's surge of greenhouse gases and drove more gradual warming. Life had more time to adapt than it does today: Fossil records show trees migrated uphill and to higher latitudes, with animals following in their wake, even as tropical corals disappeared and ecosystems wholly changed.

Past explanations for the PETM centered on methane, a greenhouse gas even more powerful than CO_2 although shorter lived. Samples of ancient plankton shells seemed to show the atmosphere during the brief hothouse was enriched in light carbon, the isotope favored by life. That suggested the carbon responsible for the warming surge originated in living things, as most methane does, rather than in the gases spewed by volcanoes, which rise from deep Earth.

At first, researchers thought a small amount of warming might have destabilized methane hydrates—seafloor deposits of methane trapped in cages of ice crystals—triggering a massive release of carbon. But the 2010 *Deepwater Horizon* oil spill in the Gulf of Mexico put a dent in that theory. Microbes simply chewed up the methane the broken well released into the ocean, suggesting seeps of seabed methane would rarely get all the way into the air. "Most modeling studies suggest you can't release enough greenhouse gases just through hydrates," says Sev Kender, a paleoceanographer at the University of Exeter.

Mudrocks on the sea floor also contain carbon that originated in living things, and magma from submarine eruptions could have heated the rocks and liberated the carbon. But in 2017, researchers analyzed plankton fossils from an ocean core and found the carbon released during the PETM was heavier than previously thought. For some, that indicated the carbon wasn't from living sources. "Given the current state of knowledge, it seems likely to be volcanism," says Marcus Gutjahr, a geochemist at GEOMAR Helmholtz Centre for Ocean Research Kiel, who led the 2017 study.

Greenland was rifting away from Europe at the time of the PETM as a mantle plume traveled under the island, priming the 180-kilometer-thick crust above to be pulled apart. Like all volcanism, the process would have released CO_2. Gernon calculated, however, that the eruptions during the rifting would have only provided one-fifth of the more than 10,000 gigatons of carbon needed to explain the PETM warming. But he knew that over the eons, CO_2 and other gases can bubble out of tectonic plates as they dive into the mantle, percolating up into the underside of thick crusts like Greenland's, and forming carbonate formations that can be stable for millions or even billions of years.

If the crust is ever pulled apart by rifting, however, the trapped carbon can spill upward and erupt as rare carbonatite lava, which contains far more CO_2 than standard lava. Indeed, such a process appears to be underway in East Africa right now, where a rift has begun to tear the horn of Africa away from the rest of the continent, says James Muirhead, a structural geologist at the University of Auckland. "At the very edge of the craton we get these carbonatite lavas," he says. "And adjacent to the craton we get high CO_2 fluxes."

Similarly, the hot spot that burned through Greenland starting 60 million years ago could have mobilized any carbonate under its crust, Gernon says. When the rifting began to open up what today is the northeastern Atlantic Ocean, "you'll have a huge amount of carbon venting."

> **"Since dinosaurs kicked the bucket, this is the biggest global warming event we have."**
>
> **Pincelli Hull**, Yale University

Evidence of the carbon-rich melt is abundant on either side of the North Atlantic rift, the tectonic division that marks the old boundary between Greenland and Europe, Gernon and his co-authors report in a study published last week in *Nature Geoscience*. In an ocean core collected in 1981, they found volcanic tuffs indicating a sharp increase in volcanism during the PETM. They also combed the literature for studies of other rocks matching the core, and found reports in East Greenland and the Faroe Islands of anomalous lavas rich in magnesium, titanium oxide, and rare earth elements—signatures of melting of carbonate rock from deep in the crust. The lavas date roughly to 56.1 million years ago, and the investigators calculate that the rifting would have produced enough of them to explain nearly all of the needed carbon emissions.

Kender says Gernon makes a compelling case, but the timing is key. The PETM lasted only several thousand years, a geological instant. Meanwhile, the volcanism has not been precisely dated. "Whether it was at the onset, in the middle, or later, we can't say yet," Kender says. Gernon's team says more precise dating from the ocean core, still unpublished, supports the idea that the lavas they're studying could be from the onset of the PETM. "I'm quietly confident the story works," Gernon says. ■

CONSERVATION BIOLOGY

In the Balkans, researchers mobilize to protect a river

Plans to dam the upper Neretva River draw concern

By **Richard Schiffman**

Five years ago, researchers from across Europe converged on a cold, fast-moving river in the highlands of Albania for a week of intensive fieldwork. Their mission: to kick off a multiyear effort to assemble a detailed ecological portrait of the Vyosa River, one of Eastern Europe's last free-flowing waterways. They hoped to draw public attention to the river's rich wildlife and persuade policymakers to protect it from a cascade of proposed dams.

Last month, that effort paid off when Albania's prime minister pledged to create a Wild River National Park that would protect some 500 kilometers of the Vyosa and its tributaries from hydropower development.

Now, scientists are hoping to replicate that success along another imperiled waterway in the region: the upper Neretva River, which is threatened by some 70 proposed dams. This week, three dozen researchers—including experts on fish, amphibians, and invertebrates—will fan out along the river's headwaters in Bosnia and Herzegovina as part of Neretva Science Week, which aims to catalog species that depend on the river.

"It is miraculous that these rivers have survived," says ecologist Ulrich Eichelmann, founder of RiverWatch, a Vienna-based group behind both the Vyosa research campaign and the Neretva Science Week, which began on 28 June. Western Balkan rivers teem with aquatic life, including the world's highest diversity of trout species, suggesting to some researchers that they are an evolutionary cradle of trout and salmon. It's an unexpected wilderness in a heavily developed continent, Eichelmann says. "In other parts of Europe, there are virtually no rivers left that have not been dammed or channeled."

Four major dams interrupt the lower reaches of the Neretva, which flows about

exceeded the total of today's known oil and gas reserves, it was slower than today's surge of greenhouse gases and drove more gradual warming. Life had more time to adapt than it does today: Fossil records show trees migrated uphill and to higher latitudes, with animals following in their wake, even as tropical corals disappeared and ecosystems wholly changed.

Past explanations for the PETM centered on methane, a greenhouse gas even more powerful than CO_2 although shorter lived. Samples of ancient plankton shells seemed to show the atmosphere during the brief hothouse was enriched in light carbon, the isotope favored by life. That suggested the carbon responsible for the warming surge originated in living things, as most methane does, rather than in the gases spewed by volcanoes, which rise from deep Earth.

At first, researchers thought a small amount of warming might have destabilized methane hydrates—seafloor deposits of methane trapped in cages of ice crystals—triggering a massive release of carbon. But the 2010 *Deepwater Horizon* oil spill in the Gulf of Mexico put a dent in that theory. Microbes simply chewed up the methane the broken well released into the ocean, suggesting seeps of seabed methane would rarely get all the way into the air. "Most modeling studies suggest you can't release enough greenhouse gases just through hydrates," says Sev Kender, a paleoceanographer at the University of Exeter.

> **"Since dinosaurs kicked the bucket, this is the biggest global warming event we have."**
>
> **Pincelli Hull**, Yale University

Mudrocks on the sea floor also contain carbon that originated in living things, and magma from submarine eruptions could have heated the rocks and liberated the carbon. But in 2017, researchers analyzed plankton fossils from an ocean core and found the carbon released during the PETM was heavier than previously thought. For some, that indicated the carbon wasn't from living sources. "Given the current state of knowledge, it seems likely to be volcanism," says Marcus Gutjahr, a geochemist at GEOMAR Helmholtz Centre for Ocean Research Kiel, who led the 2017 study.

Greenland was rifting away from Europe at the time of the PETM as a mantle plume traveled under the island, priming the 180-kilometer-thick crust above to be pulled apart. Like all volcanism, the process would have released CO_2. Gernon calculated, however, that the eruptions during the rifting would have only provided one-fifth of the more than 10,000 gigatons of carbon needed to explain the PETM warming. But he knew that over the eons, CO_2 and other gases can bubble out of tectonic plates as they dive into the mantle, percolating up into the underside of thick crusts like Greenland's, and forming carbonate formations that can be stable for millions or even billions of years.

If the crust is ever pulled apart by rifting, however, the trapped carbon can spill upward and erupt as rare carbonatite lava, which contains far more CO_2 than standard lava. Indeed, such a process appears to be underway in East Africa right now, where a rift has begun to tear the horn of Africa away from the rest of the continent, says James Muirhead, a structural geologist at the University of Auckland. "At the very edge of the craton we get these carbonatite lavas," he says. "And adjacent to the craton we get high CO_2 fluxes."

Similarly, the hot spot that burned through Greenland starting 60 million years ago could have mobilized any carbonate under its crust, Gernon says. When the rifting began to open up what today is the northeastern Atlantic Ocean, "you'll have a huge amount of carbon venting."

Evidence of the carbon-rich melt is abundant on either side of the North Atlantic rift, the tectonic division that marks the old boundary between Greenland and Europe, Gernon and his co-authors report in a study published last week in *Nature Geoscience*. In an ocean core collected in 1981, they found volcanic tuffs indicating a sharp increase in volcanism during the PETM. They also combed the literature for studies of other rocks matching the core, and found reports in East Greenland and the Faroe Islands of anomalous lavas rich in magnesium, titanium oxide, and rare earth elements—signatures of melting of carbonate rock from deep in the crust. The lavas date roughly to 56.1 million years ago, and the investigators calculate that the rifting would have produced enough of them to explain nearly all of the needed carbon emissions.

Kender says Gernon makes a compelling case, but the timing is key. The PETM lasted only several thousand years, a geological instant. Meanwhile, the volcanism has not been precisely dated. "Whether it was at the onset, in the middle, or later, we can't say yet," Kender says. Gernon's team says more precise dating from the ocean core, still unpublished, supports the idea that the lavas they're studying could be from the onset of the PETM. "I'm quietly confident the story works," Gernon says. ■

CONSERVATION BIOLOGY

In the Balkans, researchers mobilize to protect a river

Plans to dam the upper Neretva River draw concern

By **Richard Schiffman**

Five years ago, researchers from across Europe converged on a cold, fast-moving river in the highlands of Albania for a week of intensive fieldwork. Their mission: to kick off a multiyear effort to assemble a detailed ecological portrait of the Vyosa River, one of Eastern Europe's last free-flowing waterways. They hoped to draw public attention to the river's rich wildlife and persuade policymakers to protect it from a cascade of proposed dams.

Last month, that effort paid off when Albania's prime minister pledged to create a Wild River National Park that would protect some 500 kilometers of the Vyosa and its tributaries from hydropower development.

Now, scientists are hoping to replicate that success along another imperiled waterway in the region: the upper Neretva River, which is threatened by some 70 proposed dams. This week, three dozen researchers—including experts on fish, amphibians, and invertebrates—will fan out along the river's headwaters in Bosnia and Herzegovina as part of Neretva Science Week, which aims to catalog species that depend on the river.

"It is miraculous that these rivers have survived," says ecologist Ulrich Eichelmann, founder of RiverWatch, a Vienna-based group behind both the Vyosa research campaign and the Neretva Science Week, which began on 28 June. Western Balkan rivers teem with aquatic life, including the world's highest diversity of trout species, suggesting to some researchers that they are an evolutionary cradle of trout and salmon. It's an unexpected wilderness in a heavily developed continent, Eichelmann says. "In other parts of Europe, there are virtually no rivers left that have not been dammed or channeled."

Four major dams interrupt the lower reaches of the Neretva, which flows about



The rugged landscape surrounding the upper Neretva River has long helped protect it from development, but a hydropower boom poses new challenges.

225 kilometers to the Adriatic Sea in Croatia. But development largely sidestepped the river's remote upper region when it was part of the former Yugoslavia, and in recent decades war and ethnic tensions discouraged investors.

Starting around 2010, however, hydropower developers saw a promising new market in the Balkans, and the free-flowing streams beckoned. Some 3500 power dams are now either proposed or under construction in the region. Along the upper Neretva, engineers have identified dozens of sites for mostly small hydroelectric dams.

The dam-building rush is an ecological disaster in the making, says ichthyologist Steven Weiss of the University of Graz. He fears "poorly designed and poorly managed" barriers could result in "death by 1000 cuts" along the Neretva and other rivers, with fish and other organisms losing 30% to 70% of their current habitat.

During the science week, Weiss will be surveying populations of the endangered soft-mouthed trout (*Salmo obtusirostris*) and other endemic fish species. Dams would block the movements of the rare trout and fragment populations, he says. Sediment from some reservoirs would be flushed downstream, burying habitat. New lakes could become home to invasive fish, such as carp, that compete with native species.

The river's invertebrate residents will be the focus of biogeochemist Gabriel Singer of the University of Innsbruck. He and two students also plan to measure water levels of carbon dioxide and methane from the decomposition of organic matter in the streams. The data could help researchers predict how detritus trapped in stagnant reservoirs could cause production of the two potent greenhouse gases to surge. Dams are often lauded for producing clean energy, Singer notes, but "what we really should be thinking about is [their] climate warming potential."

While the other scientists catalog life aboveground, a team led by biologist Maja Zagmajster of the University of Ljubljana will be searching below the surface. The porous karst (limestone) geology of the Dinaric mountains, where the Neretva rises, has made the region a global hot spot for subterranean biodiversity, she says. Some of the resident species—such as the olm (*Proteus anguinus*), a blind cave amphibian that may live more than 100 years—inhabit caves and crevices. But others occupy the tiny water-filled spaces between grains of sand and gravel under the flowing river. Dams can pose a special threat to these "interstitial habitats," Zagmajster notes, because they alter seasonal flooding patterns and allow fine sediment to settle and choke the cavities.

Biologist Saudin Merdan, who leads the Stjepan Bolkay Center, a research institute in Bosnia, will be documenting reptiles and amphibians along the Neretva. But he also worries about the impact of hydropower development on people. Some area residents depend on the river for drinking water, whose purity could be threatened by toxic organisms that thrive in stagnant water, he notes. And, he adds, "Many don't have the money to go to the seaside on vacation, so they spend their leisure time on the rivers."

Merdan and others hope the data harvested during the science week will help catalyze support for a new national park to preserve the Neretva's headwaters. The goal, Eichelmann says, is not "safeguarding one small and scenic stretch of it, but protecting the whole network of the river, which I like to compare to a tree and its branches." ■

Going with the flow

Researchers from across Europe converged on the Neretva River on 28 June for a "science week" aimed at surveying wildlife and building the case for conservation.



Richard Schiffman is a journalist in New York City.

CREDITS: (PHOTO) VERA KNOOK; (MAP) N. DESAI/*SCIENCE*

BIOMEDICINE

Funds dwindle for NIH program for puzzling cases

Patient groups, hospitals urge continued support for the Undiagnosed Diseases Network

By **Jocelyn Kaiser**

Ten years ago, an athletic 12-year-old from Affton, Missouri, named Mitchell Herndon began to experience muscle weakness that eventually put him in a wheelchair. After years of visits to specialists failed to diagnose his neurological symptoms, he enrolled in a National Institutes of Health (NIH)-funded program that studies patients with debilitating mystery diseases. Researchers eventually found an explanation for Mitchell's condition: a mutated gene that causes certain brain cells to produce an overactive protein that leads to neuron damage.

Mitchell died 3 years ago at age 19. Since then, 14 more people have been identified with the same disease, dubbed Mitchell syndrome, and his family has started a foundation that hopes to develop a treatment.

His case is one of hundreds cracked by a highly regarded NIH program called the Undiagnosed Diseases Network (UDN) over the past 9 years. Now, however, funding is winding down for the network—and some, including patient advocates, are scrambling to persuade Congress to restore full funding or keep at least parts of it afloat.

The UDN supports teams of clinicians, geneticists, and other experts who study patients with medical cases that have confounded their doctors and specialists. It was designed to sunset 10 years after being established in 2013, and a new NIH program to carry on its work will get less than one-third as much as its current funding. With their NIH support ending in a year, seven of the network's 12 clinical sites are turning away new patients and focusing on current cases.

"The significant funding cut is a challenge," says genetic counselor Kimberly LeBlanc, director of the UDN Coordinating Center at Harvard Medical School. But she and others hope federal lawmakers will restore more UDN funding in NIH's 2023 spending bill, which is working its way through Congress.

"I want to see the UDN sites not only continue but to [also] grow and make even more discoveries and … impacts on the lives of patients and families," Michele Herndon, Mitchell's mother, recently wrote to a staffer for Senator Roy Blunt (R–MO).

The UDN emerged from an effort launched in 2008 by the NIH Clinical Center in Bethesda, Maryland, to study people with puzzling symptoms using exhaustive clinical workups and exome sequencing, a then-new tool that scours a person's protein-coding DNA for a culprit gene. The program's success led NIH's Common Fund, dedicated to new programs, to launch a national network that grew to include the NIH intramural program and 11 medical centers around the country. Funding rose from $10 million in 2013 to about $30 million for several years, then tapered to an estimated $16 million this year.

It is an undisputed success: The network has solved about 30% of nearly 1900 cases so far, finding hundreds of novel disease mutations and nearly 50 new disorders. It has published more than 175 papers and drawn a stream of positive TV and press stories.



Mitchell Herndon's puzzling neurological disorder was diagnosed by the National Institutes of Health's Undiagnosed Diseases Network.

PHOTO: MICHELE HERNDON

Like other Common Fund programs before it, the UDN needed to find a home within NIH's 27 institutes. The current plan is for the National Institute of Neurological Disorders and Stroke (NINDS) to lead a new data and coordinating center, with four other institutes helping provide an annual $5 million budget for 5 years. NIH sees the UDN "evolving into a larger, self-sustained network," a notice says.

The current sites can apply to join—but the agency plans to directly fund only the intramural site at NIH. In part, that's because genome sequencing has become a routine part of medicine that is often covered by insurance, an NIH spokesperson notes, although UDN researchers say only some insurers cover it.

The new NINDS coordinating center will fund small research grants, but scientists say this won't be enough to compensate for the network's existing research resources. Those include model organism screening centers, a metabolomics center, and research at the clinical sites such as RNA sequencing. Without these research components, the UDN will be "less robust," says clinical geneticist Vandana Shashi, who heads Duke University's UDN site.

Also troubling to UDN scientists and advocates is the fact that more than 1300 cases remain unsolved, often because scientists need to create an animal model or find another patient to confirm a suspected disease mutation. "Research is still needed to find an answer," LeBlanc says.

The NIH spokesperson says the agency is "working with Congress on other models to expand [the UDN's] impact," particularly in underserved communities where many patients can't get the basic diagnostic workup required to qualify for the UDN.

Meanwhile, some sites have cobbled together donations, parts of other NIH grants, and other support and expect to reopen in a few months. Others are still figuring out whether they can keep going.

"We hope to continue this work in any way possible, since we are passionate and committed to this mission," Shashi says. ■



Over time, the work of cheesemakers, like this one in Switzerland, has domesticated bacteria.

MICROBIOLOGY

Foodmaking microbes bear marks of domestication

Bacteria and fungi behind cheese, soy, and more share genomic traits wth domesticated plants and animals

By **Elizabeth Pennisi**

The burst of flavor from summer's first sweet corn and the proud stance of a show dog both testify to the power of domestication. But so does the microbial alchemy that turns milk into cheese, grain into bread, and soy into miso. Like the ancestors of the corn and the dog, the fungi and bacteria that drive these transformations were modified for human use. And their genomes have acquired many of the classic signatures of domestication, researchers reported in two talks last month at a meeting in Washington, D.C.

Microbes can't be "bred" in the normal sense, because unlike peas or pigs, individual microbes with desired traits can't be chosen and mated. But humans can grow microbes and select variants that best serve our purposes. The studies show the process, repeated over thousands of years, has left genetic hallmarks similar to those in domesticated plants and animals: The microbes have lost genes, evolved into new species or strains, and become unable to thrive in the wild.

The studies "are getting to the mechanisms" of how microbial domestication works, says Benjamin Wolfe, a microbiologist at Tufts University. By revealing which genes are key to microbes' prized traits—and which can be lost—the work could help further improve the organisms that fashion much of our food and drink, "especially [with] increasing interest in fermented foods," says microbial ecologist Ariane Peralta of East Carolina University.

The yeasts used in making bread have long been seen as domesticated because they have lost genetic variation and can't live in the wild. But for other microbes, scientists have been "lacking clear evidence of domestication … in part because [their] microbial communities can be hard to study," says graduate student Vincent Somerville of the University of Lausanne.

Somerville and John Gibbons, a genomicist at the University of Massachusetts, Amherst, independently focused on food fermentation, which helped early farmers and herders transform fresh produce and milk into products that can last months or years. Gibbons took a close look at the genome of *Aspergillus oryzae*, the fungus that jump-starts production of sake from rice and soy sauce and miso from soybeans.

When farmers cultivate *A. oryzae*, the fungus—a eukaryote, with its DNA enclosed in a nucleus—reproduces on its own. But when humans take a little finished sake and transfer it to a rice mash to begin fermentation anew, they also transfer cells of the fungal strains that evolved and survived best during the first round of fermentation.

Gibbons compared the genomes of scores of *A. oryzae* strains with those of their wild ancestor, *A. flavus*. Over time, he found, selection by humans had boosted *A. oryzae*'s ability to break down starches and to tolerate the alcohol produced by fermentation. "The restructuring of metabolism appears to be a hallmark of domestication in fungi," he reported last week at Microbe 2022, the annual meeting of the American Society for Microbiology. For example, domesticated *Aspergillus* strains may have up to five times more copies of a gene for metabolizing starches as their ancestor—"a brilliant way for evolution to turn up this enzyme," Wolfe says.

The genes of domesticated *A. oryzae* also display little variation, and the genome has lost some key genes, including those for toxins that would kill the yeast needed to complete fermentation—and which can make humans sick. Domestication has apparently made *A.oryzae* more human friendly, just as it bred bitter flavors out of many food plants.

Somerville reported at the meeting that he has seen much the same pattern in prokaryotes, or organisms without nuclei, including in bacteria used to make cheese. Early cheesemakers established "starter" bacterial cultures, which people in Switzerland use to make Gruyère and other cheeses. Since the 1970s, cheesemakers have banked samples of their starter cultures to evaluate their cheese and keep quality high. Somerville sequenced the genomes of more than 100 samples.

"The exciting thing from this work was having samples over time," Wolfe says. "You can see the shaping of diversity," with changes in the past 50 years hinting at the trajectory of change over past centuries.

All the samples had low genetic diversity, with just a few strains of two dominant species, Somerville reported. Those few persistent strains are probably important to cheese quality, Gibbons said. The cultures had also lost genes since the 1970s, including some needed to make certain amino acids, which are required to assemble proteins. But amino acids are costly to make—and these microbes live in protein-rich milk. "They were able to let go of a bunch of genes they didn't need," Wolfe says. Somerville also found extensive gene swapping among the microbes, one way to acquire new genes.

Putting the studies together, Gibbons concludes the genomes of "domesticated prokaryotes and microbial eukaryotes are very similar" to each other and to multicellular domesticated organisms. Peralta cautions that the analogy to crops and animals isn't perfect. Microbes can evolve far more quickly and so can more easily be "rewilded." Still, as researchers fine-tune domesticated microbes, she hopes for even better tasting sake and cheese. ■

PHOTO: DENIS BALIBOUSE/REUTERS

GLOBAL HEALTH

WHO holds off on global alarm for monkeypox

Critics blast agency for not declaring a Public Health Emergency of International Concern

By **Kai Kupferschmidt**

The World Health Organization (WHO) may have very high aspirations—"the attainment by all peoples of the highest possible level of health"—but when a new human disease begins to spread, or a known one behaves in unusual, threatening ways, it has few levers to pull. One important decision it can make, however, is declaring a Public Health Emergency of International Concern (PHEIC), a designation that gives WHO some additional powers and serves as a global wake-up call.

After a meeting behind closed doors, a WHO expert panel announced on 25 June that the rapidly growing monkeypox epidemic does not yet warrant that status—a verdict the agency has accepted. The panel's conclusion was widely criticized by virologists, epidemiologists, and public health experts, and it has triggered a new debate about the purpose of the PHEIC, an instrument created to help improve the response to international health threats. "I think they made a big mistake," says Yale School of Public Health epidemiologist Gregg Gonsalves, who advised the committee. "They punted."

WHO has previously come under fire for waiting too long to declare a PHEIC for COVID-19 and Ebola epidemics in West Africa and the Democratic Republic of the Congo. The agency's "track record is that they tend to be on the later side of things," says Jeremy Youde, a global health researcher at the University of Minnesota, Duluth. "Right now PHEICs send the message that WHO is the last institution to grasp that a newly identified outbreak is indeed a public health emergency of international concern," adds biologist Michael Worobey of the University of Arizona. "The window may already have closed on stopping the establishment of a new sexually transmitted disease worldwide, but a PHEIC has not even been declared."

The panel didn't fully explain why the monkeypox outbreak doesn't deserve PHEIC status. The International Health Regulations, a global treaty updated in 2005, define a PHEIC—pronounced as "fake"—as an "extraordinary event which is determined to constitute a public health risk to other States through the international spread of disease and to potentially require a coordinated international response." With more than 4000 monkeypox cases since May in more than 40 countries where it is not normally seen, global spread is clearly a concern, and the panel says the outbreak requires "collaborative international efforts."



People line up at a New York City pop-up clinic offering a monkeypox vaccination.

PHOTO: TAYFUN COSKUN/ANADOLU AGENCY/GETTY IMAGES

That leaves just one criterion in doubt, says Alexandra Phelan, a lawyer at Georgetown University who specializes in global health policy: whether the surge is an "extraordinary event." That determination is tricky because monkeypox is an endemic, if neglected, disease in some African countries. "It is unjust and unethical to determine an event is only extraordinary if it is now occurring in high-income countries," Phelan says. Rather than holding off on a PHEIC, however, WHO should re-examine how equitable the criteria are, she says.

Emory University virologist Boghuma Titanji notes that the question isn't just where monkeypox is spreading, but also how. The vast majority of recent cases have been in men who have sex with men, a pattern not seen before. The panel's report suggests its authors declined to recommend a PHEIC even though all criteria were met, says Clare Wenham, a global health expert at the London School of Economics. "We have seen this inconsistent application of the criteria of a PHEIC throughout its history," she says. The yes-or-no nature of the decision probably leads experts to be cautious, Youde says: "It's tricky, because it's such a blunt instrument."

How helpful a PHEIC truly is remains a matter of debate as well. The declaration obligates member countries to follow WHO's recommendations, such as sharing data on cases, and allows the agency to issue travel advice. "To date there is no scientific evidence on the effect a PHEIC has," Wenham says.

But declaring a PHEIC puts the spotlight on a pathogen, Phelan says: "I worry what waiting a few weeks before sufficiently grabbing political attention will mean for community transmission." A delay can also let an outbreak grow so big that countries are more likely to hoard vaccines and other resources. "As more countries become affected, they're less able to help countries affected early on," Phelan says. And there is no evidence that calling a PHEIC early has downsides, Youde says.

The panel acknowledged that some of its own members disagreed with the decision and said it will reconsider the issue if there is new evidence that monkeypox is spreading to other countries or new groups, or that the severity of the disease is increasing. Signs the virus is establishing new animal reservoirs outside Africa might also warrant a reconsideration.

"I think some of these criteria will be met almost immediately," Gonsalves says. The list of affected countries is growing almost daily and there have already been reports of the disease in children. The panel's report mentions a potential re-evaluation in 21 days, Gonsalves notes, "but I hope [WHO's] director-general is smart enough to convene people again in a much more timely manner." ■



FEATURES



SEA CHANGE

Can former pirates help scientists study troubled waters? *By* **David Grimm**

Laura Sánchez Alòs had been at sea for less than 20 minutes when the motorboats attacked. In late January 2019, the 24-year-old Spanish conservation biologist was on board the *Farley Mowat*, a former U.S. Coast Guard cutter, in Mexican waters in the northern Gulf of California. She was the ship's scientist, part of a campaign by the Sea Shepherd Conservation Society to study and protect the vaquita, the world's smallest porpoise and most endangered marine mammal. Since 2015, the organization had been patrolling this region—and making local fishers increasingly angry.

Sánchez Alòs was here to document the size and number of nets placed in the water to snag the 100-kilogram totoaba, a fish so highly valued for its swim bladder in traditional Chinese medicine and cuisine, it's known as the "cocaine of the sea." But the illegal nets snare vaquitas as well, and she and her colleagues were also yanking up the nets and destroying them.

Suddenly, nearly two dozen motorboats surrounded the *Mowat*. The men aboard hurled insults, calling the Sea Shepherd crew invaders and slicing their hands across their throats in an "I'll kill you" gesture. Then they began to throw rocks, iron sinkers, and Molotov cocktails. The cutter blasted the attackers with firehoses, but a large rock crashed through the ship's front window, showering glass across the bridge. A few minutes later, a Molotov cocktail set fire to the *Mowat*'s port side. The ship turned its hoses on itself to extinguish the flames, then zoomed out of the fray, pockmarked but intact. "Either they couldn't follow us anymore," Sánchez Alòs says, "or they ran out of rocks."

The incident was a turnabout for an organization used to being on the offensive. For more than 3 decades, Sea Shepherd had made a name for itself as a militant conservation group, sinking whaling vessels, throwing its own projectiles at illegal fishing boats, and generally conducting itself as a vigilante of the high seas. International courts branded the group's head, Paul Watson, a pirate and a terrorist.

Yet the Sea Shepherd Conservation Society appears to have undergone a dramatic transformation in the past 8 years. It has abandoned militant tactics in favor of new leadership and a new mission: Protect the oceans not with violence, but with science. Call it a sea change.

The group has just hired its first director of science and begun recruiting researchers from across the globe. With a dozen vessels and a $20 million annual budget, it hopes to fill an urgent scientific need: ferrying scientists, especially those from less developed nations, to understudied parts of the ocean. There they can survey endangered animals, search for new species, and even board illegal vessels to collect firsthand data on overfishing.

"We can't do what we want going forward without science driving us," says Pritam Singh, who effectively took over for Watson in 2014. "Science will be our primary mission—not an afterthought."

The pivot couldn't come at a better time for marine research. The U.S. government's scientific fleet has shrunk by one-third over the past couple of decades, according to some estimates. Time on research vessels can take years to procure—if investigators can get on at all. And overfishing, pollution, and climate change are roiling the seas in ways researchers are struggling to measure.

"There are a lot of good ocean scientists, but they're stranded on shore," says Douglas McCauley, a marine ecologist at the University of California, Santa Barbara.

Sea Shepherd's vessels could be their route to sea. But most of its ships are relatively small and only outfitted with basic

PHOTOS: (LEFT TO RIGHT) MELISSA ROMAO/SEA SHEPHERD; JA



Sea Shepherd Conservation Society ships monitor a vaquita refuge in the Gulf of California, where scientists aboard the *Sharpie* (far left), scan for the rare porpoises.

scientific equipment. And the organization itself will need to shed its past and commit to robust and transparent science, McCauley says. "It will take a rather significant about-face, but it could be really transformative," he says. "We are looking at a rough next few decades for the ocean, and we need all hands on deck."

SEA SHEPHERD TRACES its roots to 1977, when Watson was ousted from the board of Greenpeace for his controversial tactics. He formed Sea Shepherd that year and soon developed—and embraced—a reputation as a brash and militant leader. He wore a captain's uniform, holstered a knife, and flew a skull and crossbones on the ships he commanded.

For the next several decades, Sea Shepherd crews operated like vigilantes in the name of conservation. They dropped spiky steel "net rippers" onto the ocean floor to tear up trawling nets, lobbed smoke bombs onto boat decks, and threw rancid butter to taint whale meat. From 1979 to 1994, Watson claims his group sunk five whaling ships and scuttled two more.

The campaigns, chronicled in the TV series *Whale Wars*, disrupted illegal fishing, dolphin hunts, and whaling around the globe. But by the early 2010s, Interpol had declared Watson an internationally wanted fugitive and a U.S. federal court labeled his crew "pirates." Critics also charged that the group's tactics had largely backfired, bolstering pro-whaling sentiment in places like Japan and Iceland. "[Watson's] involvement in all of this is an absolute disaster," Sidney Holt, one of the main architects of the first whaling moratorium, told *The New Yorker* in 2007. "Almost everything he has been doing has had blowback for those who want to see an end to whaling."

Watson, who contends his efforts saved countless marine animals, stepped away from the Sea Shepherd Conservation Society in 2014, although he continues to direct other arms of the movement, including Amsterdam-based Sea Shepherd Global. Singh, a wealthy developer and environmentalist, and Roger Payne, a leading expert on whale behavior, had recently joined the board of the U.S.-based Sea Shepherd Conservation Society, and began to steer a new course. "By the summer of 2014, we were re-evaluating how best to accomplish our mission of protecting marine wildlife," Singh says. "It required a complete change of direction in terms of culture and approach."

The Sea Shepherd movement turned more of its focus toward overfishing and began to collaborate with foreign governments to help them monitor their waters. A major turning point came in 2015 when Sea Shepherd Global's *Bob Barker*—a former whaling ship painted with shark jaws on the bow—chased an illegal fishing vessel, *Thunder*, for 110 days and 10,000 nautical miles. The pursuit, the longest in maritime history, ended with the scuttling of the trawler and the arrest of its captain. A U.S. Navy report later called the campaign a "game changer" for the Sea Shepherd movement. "It was now operating with nation states in a legitimate protection of fisheries," the report stated. "The organization has succeeded in countering illegal fishing where private maritime security companies had hoped but failed."

The Sea Shepherd Conservation Society added more ships to its fleet and launched a campaign of what Singh calls "radical collaboration," inking agreements with a number of Latin American countries to fight poaching and overfishing. It also began its first forays into science. The group dedicated the *Martin Sheen*, a 23-meter yacht, exclusively to research, and started to conduct preliminary science missions on other vessels, including surveys of the vaquita.

Singh credits the science push to Payne. “Roger wanted to know how many species were being overfished, how many sharks were being killed for finning, and whether Sea Shepherd’s efforts were affecting these numbers,” he says. “In many countries, no one is collecting this data. We realized we had the resources to do this for the first time.”

But first, the group needed a chief scientist.

JOHN PAYNE, then an independent researcher, ran into Singh by chance last summer in Vermont. The son of Roger, he spent his early life around whales and became a marine ecologist. He had no interest in joining Sea Shepherd under Watson’s reign. “If Paul had asked, I would have said no,” he says. “Their confrontational approach wasn’t a good match for my skill set.”

Singh invited the younger Payne to his lodge in southeastern Alaska, where the two watched salmon swim up the region’s rivers. Like his father, Payne sought hard data. “I object to the way a lot of animal rights and conservation organizations are careless with facts,” says Payne, who became Sea Shepherd’s director of science in October 2021. “I wanted to help, and I wanted to make sure what we said was accurate.”

Numbers on the state of the oceans are blurry. It’s thought, for example, that global fish consumption has doubled over the past 50 years, driven largely by illegal, unregulated, and unreported fishing. Up to one out of every five fish may be thrown back dead as “bycatch,” and about one-third of the world’s fish are caught at unsustainable levels (*Science*, 11 January 2019, p. 112).

Yet those numbers—used to inform everything from government regulations to conservation efforts—are just estimates, Payne notes. A single 2009 study is still the most cited paper on the global extent of illegal fishing, he says. “The oceans are cloaked in mystery,” he says. “It’s hard to do anything about illegal fishing because we don’t know the scope of the problem.”



A beaked whale surfaces during a survey by the Sea Shepherd Conservation Society’s *Martin Sheen* off Baja California.

That’s where the Sea Shepherd Conservation Society’s navy comes in. Its dozen vessels range from a 17-meter catamaran to a 56-meter former Japanese research ship. The ships venture into remote waters others often avoid because of violence or harsh weather, such as the northern Gulf of California and parts of the Southern Ocean. Sea Shepherd groups also have unique partnerships with several countries that allow them to board fishing vessels and gather catch data. In the waters off Liberia and Peru, for example, Sea Shepherd crew, in conjunction with the countries’ coast guards, are authorized to inspect cargo holds, cataloging the size and types of fishing nets and the weight, number, and species of fish being caught.

“I don’t know anyone else who can do that,” says Ray Hilborn, an expert on fisheries at the University of Washington, Seattle. “They could survey the seas in ways no one else can.”

Sea Shepherd can also offer marine scientists a faster, cheaper route to sea. The U.S. Academic Research Fleet encompasses 18 vessels, and other federal agencies such as the National Oceanic and Atmospheric Administration (NOAA) also operate ships of various sizes used for research. But it can take up to 5 years to secure a grant to get on these boats, and costs can exceed $100,000 per day on a large research vessel. “A 2-week expedition can eat up an entire grant,” McCauley says. Sea Shepherd crews are mostly volunteer, which keeps costs down. Payne says scientists should be able to board its ships in a matter of months, not years.

The organization also promises to bring researchers to rarely frequented marine protected areas, giant swaths of ocean that restrict fishing and other activities. “Many developing countries don’t have the resources to patrol their areas, so even their own scientists have trouble getting out to these locations,” Payne says. Sea Shepherd ships already doing conservation work could bring researchers onboard to document animal populations and behavior, as well as water conditions such as temperature and salinity.

The same could go for the open ocean. “There are a lot of places in the world where we have no idea what the fishing impacts are,” says Les Watling, a biological oceanographer at the University of Hawaii, Manoa. “It would be absolutely delicious to go somewhere no one has been.”

Sea Shepherd’s preliminary scientific efforts have borne fruit. The *Martin Sheen* has been sailing to the Guadalupe Island Biosphere Reserve, which encompasses several small islands off the Pacific coast of Baja California, since 2016. On the trips, marine biologist Gustavo Cárdenas Hinojosa of Mexico’s National Commission of Natural Protected Areas studies the behavior and vocalizations of elusive beaked whales, which are vulnerable to sonar and other humanmade noise. Before he got involved with Sea Shepherd, he says, his trips were so limited, it took him 3 years just to spot 31 whales. In his first 3 weeks on the *Martin Sheen*, he saw 45.

With photographs, underwater microphones, and DNA samples from tissue

PHOTO: GUSTAVO CÁRDENAS

(obtained with darts), Cárdenas Hinojosa and his colleagues have uncovered new insights into beaked whale birthing patterns, family structure, and socialization. They have also identified a possible new species. "We simply would not be able to do this work without Sea Shepherd," he says.

A handful of other scientists have used Sea Shepherd ships to study the impact of human activity on humpback whales in the Pacific, shark behavior off the coast of Gabon, and river dolphin populations in the Amazon. Other studies have tested sperm whales in the Gulf of California for pollutants, which John Payne says "may be the biggest threat to whales post-whaling."

But these projects are just a drop in the ocean. What Payne needs now is for more scientists to climb aboard.

WHEN IT COMES TO Sea Shepherd's past, Hilborn doesn't mince words. "A bunch of lunatics conducting campaigns that were internationally criminal and a total waste of time," he says. (Watson counters that he and his colleagues have never been convicted of a felony.) But Hilborn is cautiously optimistic about the Sea Shepherd Conservation Society's change in direction. "Shedding Watson gives them a lot of opportunity for rebranding," he says. He adds that John Payne is "a good choice" as director of science. (Hilborn served on Payne's Ph.D. committee and has worked with him as a consultant.)

Still, Hilborn and others say the organization faces steep challenges if it wants to become a scientific powerhouse. One is equipment. High-powered binoculars are among Sea Shepherd's fanciest scientific devices. In contrast, traditional research vessels have dedicated scientific equipment, personnel, and labs. On the *Farley Mowat*, "a laptop is really pushing it," says Barbara Taylor, a biologist with NOAA's Southwest Fisheries Science Center who surveyed vaquitas from the ship.

Bruce Appelgate echoes those concerns. He leads ship operations at the Scripps Institution of Oceanography, which operates three major scientific vessels and a collection of smaller coastal research ships. His biggest, the 83-meter *Roger Revelle*, bristles with high-tech scientific instruments, including a deep-water mapping system, gravity meters, and long-range Doppler sonar. It can also accommodate underwater remotely operated vehicles (ROVs), hosts well-equipped laboratories, and carries a paid scientific crew to operate instruments. "Doing science right takes an extraordinary amount of work," he says.

Payne admits Sea Shepherd ships can't rival the *Revelle*. Still, he says, they "are large, stable platforms that can operate in any weather," and will support most of the types of research his organization wants to do, such as surveying biodiversity, verifying catch data, and observing marine animal behavior. Hilborn adds that Sea Shepherd could easily add more instrumentation. "It's all solvable by money."

The Sea Shepherd Conservation Society currently brings in (and spends) about $20 million per year in donations. Payne hopes to use some of that on a $100,000 echosounder to acoustically measure fish populations. "We don't have unlimited money for equipment," Singh says. "But if we needed something like an ROV, we'd try to make it work."

Hilborn is more concerned about the organization's ability to balance its scientific goals with conservation. After all, its donors and volunteers are motivated by a desire to save the whales, not measure salinity. "You need to go back to the same place over and over again for scientific surveys," he says. "You can't do that if you're chasing bad guys across the ocean."

Hilborn also worries the group won't publish results that conflict with its mission—if it finds, for example, an abundance of species in an area thought to be overfished. It could create an impartial scientific board to help design and evaluate studies, he suggests.

John Payne is currently assembling an independent board to do just that. "We are committed to full transparency," Singh says. "Your data is your data. We don't censor."

Appelgate would rather the Sea Shepherd Conservation Society used its money to fund more established scientific ventures. The organization could cover ship time for scientists, for example, or partner with Scripps as another nonprofit has done. Still, he says, if the group plunges deeply into its own research, he would judge it on its merits. "I would pay more attention to the scientific rigor than to, 'Hey, these guys used to wear pirate masks.'"

JOHN PAYNE NOW spends his days calling up scientists and mapping out his organization's research future. After reaching out to more than 100 academics around the world, he says, "there is a lot of interest in using Sea Shepherd ships for science. I think we'll be turning away a lot of people."

He aims to request research proposals for ship time by the end of the year and hopes to collaborate with hundreds of scientists within 5 years.

Interest from the developing world is especially high, he says, as academics there struggle even more for access to research vessels. What Sea Shepherd could provide in these regions is "very important," Cárdenas Hinojosa says.

Sánchez Alòs agrees. Undeterred by her run-in with the fishers, she's still trying to track vaquitas in the Gulf of California. Sea Shepherd's interactions with the locals have become less intense, she says. That may be because the organization no longer yanks nets from the water; now it just tells the Mexican navy where to find them.

Sánchez Alòs is also helping build and maintain a database of all of the information Sea Shepherd ships are collecting. "We're trying to paint a better picture of what's going on in our seas," she says. "I'm sure the science is just going to keep growing—it's what the world needs." ■



Peter Hammarstedt, a captain with Sea Shepherd Global, investigates the cargo hold of a fishing trawler off the coast of West Africa.

PHOTO: SEA SHEPHERD



Herbert *et al*. report on a new method for reconstructing atmospheric CO_2 concentrations over the past 20 million years. Their findings support a long-term covariance of atmospheric greenhouse gases and the global mean temperature.

PALEOCLIMATE

Can the Miocene climate inform the future?

Another climate reconstruction shows a correlation between temperature and CO_2

By **Anna S. von der Heydt**

Atmospheric CO_2 concentrations are rapidly rising. To reliably predict the future, researchers have looked at past climates with comparable CO_2 concentrations, dating back millions of years. Although these past climates are not perfect analogs of the future, they still provide information about how the climate system functions under increased CO_2 concentrations. Because there are no direct observations of past climate states, climate variables need to be inferred from geological data, generally referred to as proxies. On page 116 of this issue, Herbert *et al.* (*1*) present a reconstruction of the evolution of atmospheric CO_2 concentration over the past 20 million years. Their data provide a consistent explanation for the long-term evolution of global temperature and ice sheets on Earth and give a glimpse into its climate future.

The global average surface temperature is expected to rise between 1.4° and 4.4°C by the end of this century relative to the preindustrial temperature, with substantial impacts expected on every aspect of natural ecosystems and human societies (*2*). The anticipated warming depends on how Earth responds to radiative perturbations driven by greenhouse gases in the atmosphere, of which CO_2 plays the most prominent role. Quantitative comparisons of future projected climates between different geological epochs have revealed the Pliocene climate [~3 million years ago (Ma)] as the most similar to the projected climate under the RCP4.5 scenario, which is a likely greenhouse gas trajectory according to the Intergovernmental Panel on Climate Change (IPCC). (RCP stands for representative concentration pathway, with the number representing the level of radiative forcing in the atmosphere by 2100, measured as watts per square meter.) Under the less optimistic RCP8.5 scenario, the world may then stabilize in an Eocene-like climate circa 50 Ma (*3*). However, during the Eocene, polar ice sheets were absent, and geographical boundary conditions were substantially different from those today, which makes direct comparison with future projected climates more difficult.

A more recent warm climate episode in the past is the Miocene, from roughly 23 to 5.3 Ma. One puzzling element of Miocene climate is that its global temperature seemed to be higher than what one would expect given its atmospheric CO_2 concentration (*4*). Proxies suggest global sea surface temperature during the Miocene Climatic Optimum [(MCO), about 15 Ma] to be 11.5°C warmer than that of the preindustrial period, although estimates of atmospheric CO_2 suggest relatively low levels of 450 to 550 parts per million (ppm) (*5*).

Thus far, climate models have failed to reconcile such warm MCO temperatures without assuming atmospheric CO_2 concentrations in excess of 850 ppm (*5*). Herbert *et al.* present newly reconstructed CO_2 concentrations of 500 to 1100 ppm during much of the early Miocene, including the MCO, suggesting that the CO_2 amounts in climate models were not unrealistically high. Most notably, their reconstruction considers the rate at which oceanic crust is produced to be directly related to tectonic degassing rates of CO_2. In previous studies, tectonic degassing rates were mostly assumed to be constant over time, but Herbert *et al.* suggest that tectonic degassing was substantially higher up until the MCO, after which degassing rates declined along with global cooling. The authors arrive at an even warmer estimate of MCO global mean surface temperatures by using organic proxies from higher-latitude sites only. They carefully chose the proxies because during warm climates, organic proxies can become saturated in tropical and mid-latitude sites and can therefore lead to unreliable results.

Hebert *et al.*'s reconstructed atmospheric CO_2 evolution over the past 20 million years tracks with estimated global mean surface temperatures. Therefore, the Miocene climate appears less like an exception to the rule (*4*). With their model for the Miocene providing a temperature and global ice volume evolution that follows the atmospheric CO_2 concentration, different episodes in the Miocene may provide climate states with similarity to future projected climate states. The authors used a carbon budget model to explore how decreasing the tectonic degassing source of CO_2 could lead to declining CO_2 concentrations. Over long time scales, the degassing CO_2 source is balanced by a temperature-dependent sink resulting from the weathering of silicate rocks, extracting CO_2 from the atmosphere (*6*). Because temperature depends on atmospheric CO_2, two parameters enter the carbon budget model—Earth system sensitivity (ESS) and weatherability. ESS relates global temperature and atmospheric CO_2 on long time scales. It differs from a parameter that is more widely used for the future called equilibrium climate sensitivity (ECS). ESS includes slower processes, such as tectonically driven land-ocean distribution, vegetation cover, and land-ice distribution (*7*, *8*), whereas ECS quantifies only the faster responses to CO_2 doubling. By varying the ESS and weatherability within plausible ranges, Herbert *et al.* arrived at their CO_2 evolution, which includes uncertainty estimates that mostly stem from uncertainty in ESS.

Institute for Marine and Atmospheric Research, Utrecht University, Utrecht, Netherlands. Email: a.s.vonderheydt@uu.nl

ILLUSTRATION: MAURICIO ANTON/SCIENCE SOURCE

Despite extensive research, both ECS and ESS are subject to considerable uncertainty (*9*). Improving the understanding of climate sensitivity and its evolution on long time scales—e.g., throughout the past 65 million years—is therefore an important challenge for the field of paleoclimate studies (*10*, *11*). The evolution of atmospheric CO_2 and temperature throughout the past 20 million years suggests that the Miocene climate is less exceptional than previously thought. Global temperature, land and sea ice volume, and atmospheric CO_2 seem to covary on long time scales in ways that are not too different before and after the Miocene. Declining CO_2 concentrations over the Cenozoic period reached a first threshold at the end of the Eocene (34 Ma), crossing a tipping point that resulted in a large ice cap on Antarctica. Although this CO_2 threshold remains subject to considerable (model-based) uncertainty, the previously reconstructed low Miocene CO_2 amounts were at odds with the observation that the East Antarctic Ice Sheet largely disappeared during the MCO.

Within the paleoclimate community, the mounting Miocene proxy data continue to drive the need for a coordinated effort to compare different Miocene models (*12*). The Miocene climate could also provide information and understanding of nonlinear or tipping point behavior of specific climate subsystems—such as polar ice caps, the global ocean circulation, or tropical rainforests—relevant for future climate change (*13*). A better understanding of the Miocene climate will provide important insights for understanding the future climate of Earth. ■

REFERENCES AND NOTES

1. T. D. Herbert *et al*., *Science* **377**, 116 (2022).
2. Intergovernmental Panel on Climate Change, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021).
3. K. D. Burke *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13288 (2018).
4. J. E. Tierney *et al*., *Science* **370**, eaay3701 (2020).
5. N. J. Burls *et al*., *Paleoceanogr. Paleocimatol*. **36**, e2020PA004054 (2021).
6. J. C. G. Walker *et al*., *J. Geophys. Res*. **86**, 9776 (1981).
7. PALAEOSENS Project Members, *Nature* **491**, 683 (2012).
8. D. J. Lunt *et al*., *Nat. Geosci*. **3**, 60 (2010).
9. S. C. Sherwood *et al*., *Rev. Geophys*. **58**, e2019RG000678 (2020).
10. A. S. von der Heydt *et al*., *Curr. Clim. Change Rep*. **2**, 148 (2016).
11. A. M. Haywood *et al*., *Earth Syst. Environ*. **3**, 1 (2019).
12. DeepMIP-Miocene, The Deep-Time Model Intercomparison Project; www.deepmip.org/deepmip-miocene/.
13. H. J. Schellnhuber *et al*., *Nat. Clim. Chang*. **6**, 649 (2016).

ACKNOWLEDGMENTS

A.S.v.d.H. acknowledges support by the program of the Netherlands Earth System Science Centre (NESSC), financially supported by the Ministry of Education, Culture and Science (OCW) and the TiPES project (TiPES contribution no. 161), which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 820970.

10.1126/science.abq6542

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

NEUROSCIENCE

Sleeping off stress

Social defeat activates midbrain cells, promoting sleep and reducing anxiety in mice

By **Marian Joëls**[1,2] *and* **E. Ronald de Kloet**[3]

When a mouse is placed in the cage of an aggressive conspecific (the "resident"), usually the intruder is quickly defeated. This experience of social defeat causes stress to the intruder. When social defeat takes place during the inactive phase of the day, most of the defeated animals subsequently fall asleep. On page 63 of this issue, Yu *et al*. (*1*) show that a small group of cells in the ventral tegmental area (VTA), projecting to the lateral hypothalamus, is activated by social defeat. This pathway is key to restorative sleep after social defeat, preventing a lasting state of anxiety.

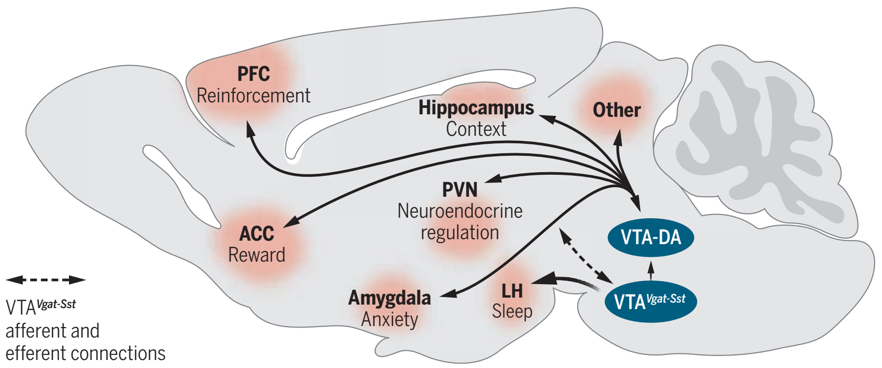
These VTA cells are γ-aminobutyric acid (GABA) neurons, and not the classical dopamine neurons that regulate reward processing and reinforcement learning (*2*), social competence (*3*), motivation, and stress coping (*4*). Yu *et al*. find that after social defeat in mice, input from brain areas such as the paraventricular nucleus of the hypothalamus, the lateral preoptic area, and periaqueductal gray activates 15 to 20% of all VTA GABAergic cells that express the vesicular GABA transporter (VTA^{Vgat}). These are particularly somatostatin-expressing cells ($VTA^{Vgat\text{-}Sst}$) and not parvalbumin-expressing cells. The $VTA^{Vgat\text{-}Sst}$ cells show increased activity up to several hours after social defeat and project to the lateral hypothalamus. Activation of this pathway is both necessary and sufficient to promote rapid eye movement (REM) and non-REM sleep onset, as well as sleep duration, after social defeat. The authors found that sleep after social defeat normalizes raised plasma corticosterone concentrations and keeps anxiety at bay. Conversely, sleep deprivation (or any interference with $VTA^{Vgat\text{-}Sst}$ cell-mediated activation of the lateral hypothalamus) sustains stress-induced elevation of corticosterone concentrations and anxiety-like behavior.

Exactly how sleep is induced after social defeat remains unresolved, as is the nature of the cells to which the $VTA^{Vgat\text{-}Sst}$ cells project. It might seem straightforward that these VTA interneurons project onto GABAergic cells in the lateral hypothalamus, which are known to locally inhibit orexin-producing cells (*5*). Orexin is a neuropeptide that, among other things, is important for arousal and wakefulness. By inhibiting orexin-producing cells, faster sleep onset, longer sleep duration, and reduction of anxiety would be predicted. However, attempts by Yu *et al*. to demonstrate a role of orexin blockade on, for example, anxiety were inconclusive.

The relationship between social defeat–induced sleep, restoration of plasma corticosterone concentrations, and prevention of

Social defeat activates interconnected brain systems

Somatostatin (Sst)-expressing γ-aminobutyric acid (GABA) neurons in the mouse ventral tegmental area ($VTA^{Vgat\text{-}Sst}$) that project to the lateral hypothalamus (LH) are necessary and sufficient to promote sleep and attenuate anxiety after social defeat. Social defeat also activates other pathways, connecting the VTA mesocorticolimbic dopaminergic (DA) system to, e.g., the amygdala, nucleus accumbens (ACC), prefrontal cortex (PFC), paraventricular nucleus (PVN), and hippocampus. These systems are often interconnected, processing various elements intrinsic to a social defeat experience.



Despite extensive research, both ECS and ESS are subject to considerable uncertainty (*9*). Improving the understanding of climate sensitivity and its evolution on long time scales—e.g., throughout the past 65 million years—is therefore an important challenge for the field of paleoclimate studies (*10*, *11*). The evolution of atmospheric CO_2 and temperature throughout the past 20 million years suggests that the Miocene climate is less exceptional than previously thought. Global temperature, land and sea ice volume, and atmospheric CO_2 seem to covary on long time scales in ways that are not too different before and after the Miocene. Declining CO_2 concentrations over the Cenozoic period reached a first threshold at the end of the Eocene (34 Ma), crossing a tipping point that resulted in a large ice cap on Antarctica. Although this CO_2 threshold remains subject to considerable (model-based) uncertainty, the previously reconstructed low Miocene CO_2 amounts were at odds with the observation that the East Antarctic Ice Sheet largely disappeared during the MCO.

Within the paleoclimate community, the mounting Miocene proxy data continue to drive the need for a coordinated effort to compare different Miocene models (*12*). The Miocene climate could also provide information and understanding of nonlinear or tipping point behavior of specific climate subsystems—such as polar ice caps, the global ocean circulation, or tropical rainforests—relevant for future climate change (*13*). A better understanding of the Miocene climate will provide important insights for understanding the future climate of Earth. ■

REFERENCES AND NOTES

1. T. D. Herbert *et al.*, *Science* **377**, 116 (2022).
2. Intergovernmental Panel on Climate Change, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte *et al.*, Eds. (Cambridge Univ. Press, 2021).
3. K. D. Burke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13288 (2018).
4. J. E. Tierney *et al.*, *Science* **370**, eaay3701 (2020).
5. N. J. Burls *et al.*, *Paleoceanogr. Paleoclimatol.* **36**, e2020PA004054 (2021).
6. J. C. G. Walker *et al.*, *J. Geophys. Res.* **86**, 9776 (1981).
7. PALAEOSENS Project Members, *Nature* **491**, 683 (2012).
8. D. J. Lunt *et al.*, *Nat. Geosci.* **3**, 60 (2010).
9. S. C. Sherwood *et al.*, *Rev. Geophys.* **58**, e2019RG000678 (2020).
10. A. S. von der Heydt *et al.*, *Curr. Clim. Change Rep.* **2**, 148 (2016).
11. A. M. Haywood *et al.*, *Earth Syst. Environ.* **3**, 1 (2019).
12. DeepMIP-Miocene, The Deep-Time Model Intercomparison Project; www.deepmip.org/deepmip-miocene/.
13. H. J. Schellnhuber *et al.*, *Nat. Clim. Chang.* **6**, 649 (2016).

ACKNOWLEDGMENTS

A.S.v.d.H. acknowledges support by the program of the Netherlands Earth System Science Centre (NESSC), financially supported by the Ministry of Education, Culture and Science (OCW) and the TiPES project (TiPES contribution no. 161), which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 820970.

10.1126/science.abq6542

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

NEUROSCIENCE

Sleeping off stress

Social defeat activates midbrain cells, promoting sleep and reducing anxiety in mice

By **Marian Joëls**[1,2] *and* **E. Ronald de Kloet**[3]

When a mouse is placed in the cage of an aggressive conspecific (the "resident"), usually the intruder is quickly defeated. This experience of social defeat causes stress to the intruder. When social defeat takes place during the inactive phase of the day, most of the defeated animals subsequently fall asleep. On page 63 of this issue, Yu *et al.* (*1*) show that a small group of cells in the ventral tegmental area (VTA), projecting to the lateral hypothalamus, is activated by social defeat. This pathway is key to restorative sleep after social defeat, preventing a lasting state of anxiety.

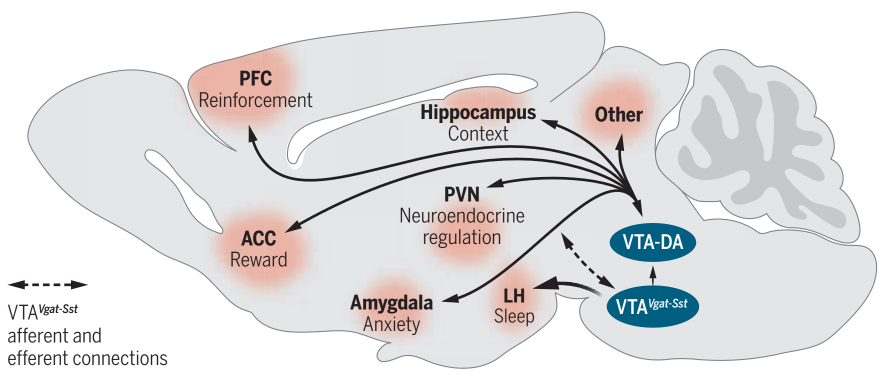
These VTA cells are γ-aminobutyric acid (GABA) neurons, and not the classical dopamine neurons that regulate reward processing and reinforcement learning (*2*), social competence (*3*), motivation, and stress coping (*4*). Yu *et al.* find that after social defeat in mice, input from brain areas such as the paraventricular nucleus of the hypothalamus, the lateral preoptic area, and periaqueductal gray activates 15 to 20% of all VTA GABAergic cells that express the vesicular GABA transporter (VTA^{Vgat}). These are particularly somatostatin-expressing cells ($VTA^{Vgat\text{-}Sst}$) and not parvalbumin-expressing cells. The $VTA^{Vgat\text{-}Sst}$ cells show increased activity up to several hours after social defeat and project to the lateral hypothalamus. Activation of this pathway is both necessary and sufficient to promote rapid eye movement (REM) and non-REM sleep onset, as well as sleep duration, after social defeat. The authors found that sleep after social defeat normalizes raised plasma corticosterone concentrations and keeps anxiety at bay. Conversely, sleep deprivation (or any interference with $VTA^{Vgat\text{-}Sst}$ cell-mediated activation of the lateral hypothalamus) sustains stress-induced elevation of corticosterone concentrations and anxiety-like behavior.

Exactly how sleep is induced after social defeat remains unresolved, as is the nature of the cells to which the $VTA^{Vgat\text{-}Sst}$ cells project. It might seem straightforward that these VTA interneurons project onto GABAergic cells in the lateral hypothalamus, which are known to locally inhibit orexin-producing cells (*5*). Orexin is a neuropeptide that, among other things, is important for arousal and wakefulness. By inhibiting orexin-producing cells, faster sleep onset, longer sleep duration, and reduction of anxiety would be predicted. However, attempts by Yu *et al.* to demonstrate a role of orexin blockade on, for example, anxiety were inconclusive.

The relationship between social defeat–induced sleep, restoration of plasma corticosterone concentrations, and prevention of

Social defeat activates interconnected brain systems

Somatostatin (Sst)-expressing γ-aminobutyric acid (GABA) neurons in the mouse ventral tegmental area ($VTA^{Vgat\text{-}Sst}$) that project to the lateral hypothalamus (LH) are necessary and sufficient to promote sleep and attenuate anxiety after social defeat. Social defeat also activates other pathways, connecting the VTA mesocorticolimbic dopaminergic (DA) system to, e.g., the amygdala, nucleus accumbens (ACC), prefrontal cortex (PFC), paraventricular nucleus (PVN), and hippocampus. These systems are often interconnected, processing various elements intrinsic to a social defeat experience.



anxiety, as reported by Yu *et al.*, is not equivocal: Are these phenomena causally linked or do they occur in parallel? A causal relationship is suggested by the observation that VTA$^{Vgat\text{-}Sst}$ stimulation during artificial sleep deprivation after social defeat cannot restrain anxiety, nor (fully) normalize stress hormone concentrations in plasma. Yet, chemogenetic stimulation of VTA$^{Vgat\text{-}Sst}$ cells (in the absence of social defeat) increases sleep but does not affect concentrations of the stress hormones corticotropin-releasing factor (CRF) and corticosterone. This suggests that parallel pathways might regulate sleep and anxiety, which is also supported by recent studies showing the importance of a reciprocal loop between the VTA and parts of the amygdala in social defeat–induced anxiety (*6, 7*). The parallel pathways of VTA–lateral hypothalamus to regulate sleep and VTA-amygdala to regulate anxiety are reminiscent of the observation that glucocorticoid receptor deletion in prefrontal dopaminoceptive neurons decreases social avoidance but not anxiety and fear memories after mice were subjected to 10 days of social defeat (*8*).

It could be hypothesized that sleep induced by the hypothalamus and activation of the hypothalamic–pituitary–adrenal (HPA) axis (which eventually causes a rise in circulating corticosterone concentration) may affect social aversion, and through a parallel pathway involving limbic regions, reduce anxiety (*9, 10*). Thus, during sleep, events experienced while awake are replayed, binding elements of the experience in various parts of the brain and forming a memory of the situation (*11*). Such memory-enhancing effects of sleep are particularly strong for locations with an aversive connotation, as in the case of a social defeat experience. Eventually this results in the formation of gist-like memories. REM sleep is proposed to depotentiate amygdala reactivation (which is associated with anxiety), tuning down amygdala inputs to the hippocampus (*12*). This would allow contextual elements to be stored, freed from the anxious context. Overall, interconnected parallel loops appear to be involved in the processing of various elements that are intrinsic to a social defeat experience (see the figure).

Sleep and stress entertain a complex courtship. Intuitively, falling asleep after experiencing stress may not sound logical: Almost all mood and anxiety disorders are characterized by sleep abnormalities, which in turn exacerbate the susceptibility to new episodes and progression of such disorders (*12, 13*). However, there may be two important differences between such pathological conditions and the one described by Yu *et al.* Single stressors may have different consequences than the wear and tear caused by a chronic state of (unpredictable) stress. Furthermore, the induction of sleep may be specific for stressful situations such as social defeat that have little ambiguity, whereas stressors with a more uncertain outcome may affect sleep negatively. Even though social defeat is a negative live event, it is relatively straightforward: The defeated mouse knows exactly what the consequence is—i.e., a subordinate position.

Not all individuals may respond to social defeat with a bout of sleep. Recent findings showed that social stress promotes sleep-like inactivity in mice, with a large degree of variation among experimental animals (*14*). This variation is also noticeable in the experiments by Yu *et al.*: About half of the animals showed a strong increase in REM sleep duration, whereas the remaining mice slept the same amount as controls. This requires further investigation in larger groups of mice (*15*), which would also be helpful to probe the robustness of the current explorative study. The source of interindividual variation could arise from the animal's sex (the social defeat model is optimized for male mice), and genetic predisposition, and be modulated by life events, affecting the sleep pattern, personality, stress responsivity, ability to contextualize, and so on—or any combination of these. Knowing the essential steps in the brain may help steer future interventions in rodents and perhaps even humans after stressful experiences, be it through cognitive therapy or pharmacotherapy or maybe, one day, genetic interference. ■

[1]Department of Translational Neuroscience, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [2]University Medical Center Groningen, University of Groningen, Groningen, Netherlands. [3]Leiden Academic Center for Drug Research, Leiden University Medical Center, Leiden University, Leiden, Netherlands. Email: m.joels@umcg.nl

REFERENCES AND NOTES

1. X. Yu *et al.*, *Science* **377**, 63 (2022).
2. W. Schultz, P. Dayan, P. R. Montague, *Science* **275**, 1593 (1997).
3. S. Ghosal, C. Sandi, M. A. van der Kooij, *Neuropharmacology* **159**, 107498 (2019).
4. C. Bouarab, B. Thompson, A. M. Polter, *Front Neural Circuits* **13**, 78 (2019).
5. L. L. Ferrari *et al.*, *J. Neurosci*. **38**, 1588 (2018).
6. N. Dedic *et al.*, *Front. Neurosci*. **13**, 986 (2019).
7. C. Morel *et al.*, *Nat. Commun.* **13**, 1532 (2022).
8. J. Barik *et al.*, *Science* **339**, 332 (2013).
9. P. Belujon, A. A. Grace, *Int. J. Neuropsychopharmacol*. **20**, 1036 (2017).
10. E. H. Douma, E. R. de Kloet, *Neurosci. Biobehav. Rev*. **108**, 48 (2020).
11. J. G. Klinzing, N. Niethard, J. Born, *Nat. Neurosci.* **22**, 1598 (2019).
12. A. N. Goldstein, M. P. Walker, *Annu. Rev. Clin. Psychol*. **10**, 679 (2014).
13. C. Baglioni *et al.*, *Psychol. Bull*. **142**, 969 (2016).
14. M. Nagai, H. Nagai, C. Numa, T. Furuyashiki, *Sci. Rep*. **10**, 19800 (2020).
15. V. Bonapersona, H. Hoijtink, R. A. Sarabdjitsingh, M. Joëls, *Nat. Neurosci*. **24**, 470 (2021).

10.1126/science.adc9782

HEALTH TECHNOLOGY

Cooling the pain

A miniaturized, flexible cooling device can be used for precise analgesia

By **Shan Jiang**[1,2] *and* **Guosong Hong**[1,2]

As one of the most critical health problems, pain afflicts one in five adults in the world (*1*). Despite the efficacy of opioids in treating pain, opioid use disorder and overdose have motivated the development of nonopioid alternatives, such as other analgesics (*2*), electrical stimulation (*3*), and acupuncture (*4*) for pain management. Among these alternatives, analgesic nerve cooling offers an effective and reversible strategy to alleviate pain (*5*). On page 109 of this issue, Reeder *et al.* (*6*) report a miniaturized and implantable cooling system that integrates state-of-the-art microfluidic and flexible electronic technologies in a biodegradable platform for localized temperature control and precise pain relief.

Analgesic nerve cooling leverages the temperature dependence of chemical reaction rates to effectively reduce the metabolic, electrogenic, and ionic activity in the neural tissue at lower temperatures (*7*). A moderate temperature decrease in nerves to 15°C has been reported to block the transmission of compound nerve action potentials, whereas a complete nerve conduction block can be achieved at 5°C (*8*). The risk of inducing nerve damage at an even lower temperature calls for precise control of local cooling. Conventional nerve-cooling methods rely on precooled liquids, such as methanol, delivered through a metal or silicone loop or a thermoelectric device (*9*). These interfaces are constrained by their bulky and rigid structures, nonspecific cooling, and high-power requirement, thus limiting their chronic application in the peripheral nervous system.

An implantable cooling device with on-demand local analgesia will be a game changer for long-term pain management. Reeder *et al.* capitalized on their expertise in advanced flexible microsystems (*10*) to create a microfluidic cooling device with real-time temperature feedback. Compared

[1]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA. [2]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA 94305, USA. Email: guosongh@stanford.edu

anxiety, as reported by Yu *et al.*, is not equivocal: Are these phenomena causally linked or do they occur in parallel? A causal relationship is suggested by the observation that $VTA^{Vgat\text{-}Sst}$ stimulation during artificial sleep deprivation after social defeat cannot restrain anxiety, nor (fully) normalize stress hormone concentrations in plasma. Yet, chemogenetic stimulation of $VTA^{Vgat\text{-}Sst}$ cells (in the absence of social defeat) increases sleep but does not affect concentrations of the stress hormones corticotropin-releasing factor (CRF) and corticosterone. This suggests that parallel pathways might regulate sleep and anxiety, which is also supported by recent studies showing the importance of a reciprocal loop between the VTA and parts of the amygdala in social defeat–induced anxiety (*6, 7*). The parallel pathways of VTA–lateral hypothalamus to regulate sleep and VTA-amygdala to regulate anxiety are reminiscent of the observation that glucocorticoid receptor deletion in prefrontal dopaminoceptive neurons decreases social avoidance but not anxiety and fear memories after mice were subjected to 10 days of social defeat (*8*).

It could be hypothesized that sleep induced by the hypothalamus and activation of the hypothalamic–pituitary–adrenal (HPA) axis (which eventually causes a rise in circulating corticosterone concentration) may affect social aversion, and through a parallel pathway involving limbic regions, reduce anxiety (*9, 10*). Thus, during sleep, events experienced while awake are replayed, binding elements of the experience in various parts of the brain and forming a memory of the situation (*11*). Such memory-enhancing effects of sleep are particularly strong for locations with an aversive connotation, as in the case of a social defeat experience. Eventually this results in the formation of gist-like memories. REM sleep is proposed to depotentiate amygdala reactivation (which is associated with anxiety), tuning down amygdala inputs to the hippocampus (*12*). This would allow contextual elements to be stored, freed from the anxious context. Overall, interconnected parallel loops appear to be involved in the processing of various elements that are intrinsic to a social defeat experience (see the figure).

Sleep and stress entertain a complex courtship. Intuitively, falling asleep after experiencing stress may not sound logical: Almost all mood and anxiety disorders are characterized by sleep abnormalities, which in turn exacerbate the susceptibility to new episodes and progression of such disorders (*12, 13*). However, there may be two important differences between such pathological conditions and the one described by Yu *et al.* Single stressors may have different consequences than the wear and tear caused by a chronic state of (unpredictable) stress. Furthermore, the induction of sleep may be specific for stressful situations such as social defeat that have little ambiguity, whereas stressors with a more uncertain outcome may affect sleep negatively. Even though social defeat is a negative live event, it is relatively straightforward: The defeated mouse knows exactly what the consequence is—i.e., a subordinate position.

Not all individuals may respond to social defeat with a bout of sleep. Recent findings showed that social stress promotes sleep-like inactivity in mice, with a large degree of variation among experimental animals (*14*). This variation is also noticeable in the experiments by Yu *et al.*: About half of the animals showed a strong increase in REM sleep duration, whereas the remaining mice slept the same amount as controls. This requires further investigation in larger groups of mice (*15*), which would also be helpful to probe the robustness of the current explorative study. The source of interindividual variation could arise from the animal's sex (the social defeat model is optimized for male mice), and genetic predisposition, and be modulated by life events, affecting the sleep pattern, personality, stress responsivity, ability to contextualize, and so on—or any combination of these. Knowing the essential steps in the brain may help steer future interventions in rodents and perhaps even humans after stressful experiences, be it through cognitive therapy or pharmacotherapy or maybe, one day, genetic interference. ■

[1]Department of Translational Neuroscience, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [2]University Medical Center Groningen, University of Groningen, Groningen, Netherlands. [3]Leiden Academic Center for Drug Research, Leiden University Medical Center, Leiden University, Leiden, Netherlands. Email: m.joels@umcg.nl

REFERENCES AND NOTES

1. X. Yu *et al.*, *Science* **377**, 63 (2022).
2. W. Schultz, P. Dayan, P. R. Montague, *Science* **275**, 1593 (1997).
3. S. Ghosal, C. Sandi, M.A. van der Kooij, *Neuropharmacology* **159**, 107498 (2019).
4. C. Bouarab, B. Thompson, A. M. Polter, *Front Neural Circuits* **13**, 78 (2019).
5. L. L. Ferrari *et al.*, *J. Neurosci*. **38**, 1588 (2018).
6. N. Dedic *et al.*, *Front. Neurosci*. **13**, 986 (2019).
7. C. Morel *et al.*, *Nat. Commun.* **13**, 1532 (2022).
8. J. Barik *et al.*, *Science* **339**, 332 (2013).
9. P. Belujon, A. A. Grace, *Int. J. Neuropsychopharmacol*. **20**, 1036 (2017).
10. E. H. Douma, E. R. de Kloet, *Neurosci. Biobehav. Rev*. **108**, 48 (2020).
11. J. G. Klinzing, N. Niethard, J. Born, *Nat. Neurosci.* **22**, 1598 (2019).
12. A. N. Goldstein, M. P. Walker, *Annu. Rev. Clin. Psychol*. **10**, 679 (2014).
13. C. Baglioni *et al.*, *Psychol. Bull*. **142**, 969 (2016).
14. M. Nagai, H. Nagai, C. Numa, T. Furuyashiki, *Sci. Rep*. **10**, 19800 (2020).
15. V. Bonapersona, H. Hoijtink, R. A. Sarabdjitsingh, M. Joëls, *Nat. Neurosci*. **24**, 470 (2021).

10.1126/science.adc9782

HEALTH TECHNOLOGY

Cooling the pain

A miniaturized, flexible cooling device can be used for precise analgesia

By **Shan Jiang**[1,2] *and* **Guosong Hong**[1,2]

As one of the most critical health problems, pain afflicts one in five adults in the world (*1*). Despite the efficacy of opioids in treating pain, opioid use disorder and overdose have motivated the development of nonopioid alternatives, such as other analgesics (*2*), electrical stimulation (*3*), and acupuncture (*4*) for pain management. Among these alternatives, analgesic nerve cooling offers an effective and reversible strategy to alleviate pain (*5*). On page 109 of this issue, Reeder *et al.* (*6*) report a miniaturized and implantable cooling system that integrates state-of-the-art microfluidic and flexible electronic technologies in a biodegradable platform for localized temperature control and precise pain relief.

Analgesic nerve cooling leverages the temperature dependence of chemical reaction rates to effectively reduce the metabolic, electrogenic, and ionic activity in the neural tissue at lower temperatures (*7*). A moderate temperature decrease in nerves to 15°C has been reported to block the transmission of compound nerve action potentials, whereas a complete nerve conduction block can be achieved at 5°C (*8*). The risk of inducing nerve damage at an even lower temperature calls for precise control of local cooling. Conventional nerve-cooling methods rely on precooled liquids, such as methanol, delivered through a metal or silicone loop or a thermoelectric device (*9*). These interfaces are constrained by their bulky and rigid structures, nonspecific cooling, and high-power requirement, thus limiting their chronic application in the peripheral nervous system.

An implantable cooling device with on-demand local analgesia will be a game changer for long-term pain management. Reeder *et al.* capitalized on their expertise in advanced flexible microsystems (*10*) to create a microfluidic cooling device with real-time temperature feedback. Compared

[1]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA. [2]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA 94305, USA. Email: guosongh@stanford.edu

with other approaches to achieving nerve blocks, their device outperforms conventional in vivo cooling methods for several reasons (see the figure). Composed of serpentine architectures and elastomeric materials with similar elasticity to those of peripheral nerves, the device is soft, flexible, and stretchable. Because of its tissue-like mechanical properties, this device readily wraps around the nerve like a cuff electrode, forming an intimate interface to facilitate effective heat transfer. The device is also bioresorbable; it dissolves in physiological fluids over months and yields a minimal inflammatory response in the tissue. The gradual biodegradation of the device is attributed to a bioresorbable elastomer known as poly(octanediol citrate), which is used to build the microfluidic system. In addition, the device leverages an evaporative mechanism to produce rapid cooling, in contrast to conventional precooled liquids with slower temperature reduction kinetics. Because the spatial spread of heat diffusion scales with time, more rapid cooling, along with the low heat capacity of exhaust gas, yields greater spatial confinement and more localized analgesia down to the millimeter scale. Lastly, this device incorporates a temperature sensor in an electronic layer alongside the microfluidic system, which provides real-time feedback for precise control of local temperatures.

In achieving spatiotemporally precise cooling, Reeder *et al.* demonstrate the power of engineering advances in addressing unmet biomedical needs. The evaporative cooling scheme has been applied in the thermal management of high-density electronics (*11*). However, its application in pain management in living organisms has so far been limited. This is largely because of the shape mismatch at the thermal interface because these devices usually come in the form of rigid parallel plates (*12*), making them less than ideal for accompanying nerves, which, by contrast, are soft and geometrically complex.

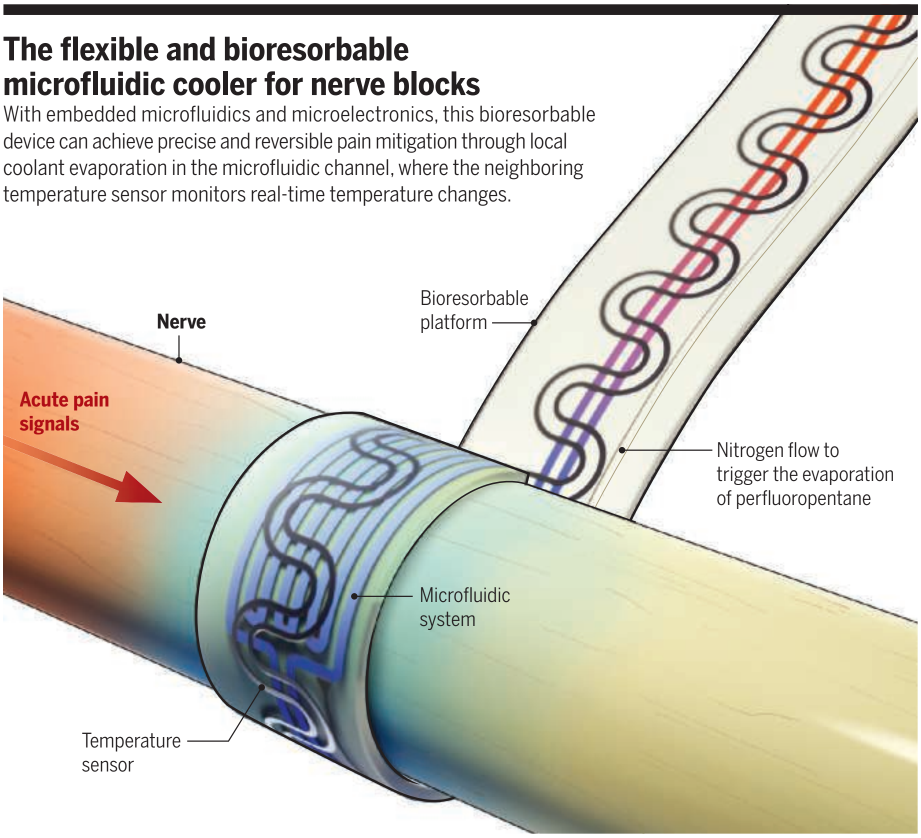
Reeder *et al.* applied evaporative cooling in a soft and stretchable microfluidic platform, which has previously been used for drug delivery in the nervous system. The bioinertness of perfluoropentane and its clinically approved use as an intravenous ultrasound contrast agent justify its use as a biocompatible liquid coolant in vivo. By taking advantage of the low boiling point and high enthalpy of vaporization of perfluoropentane (*13*) alongside the use of dry nitrogen gas, the device created by Reeder *et al.* achieved a maximum cooling rate of 3°C/s.

When tested on rats, their device produced highly localized cooling and enabled efficacious and reversible conduction blocks of rat sciatic nerves, as observed by electromyography, compound nerve action potential, and muscle-force measurements. Furthermore, the stretchable cooling device achieved analgesia in freely moving rats with neuropathic pain over several weeks after implantation, as demonstrated in von Frey anesthesiometer measurements and histological analyses. This stable interface, along with its degradation over time, enables on-demand pain management with long-term utility while obviating the need for surgical extraction after the period of intervention.

Besides the demonstrated strengths of the miniaturized, flexible cooling device for pain mitigation, the technology presents further opportunities for neuroscience research and neurological practice. Specifically, cortical cooling has been used as a nongenetic means to silence neural activities and investigate circuit dynamics in the brain (*14*). The localized, reversible cooling effect represents an attractive method for manipulating neural activities with high spatiotemporal resolutions, chronic utility, and minimal invasiveness in nonhuman primates and even humans. In addition, this technology offers a nonopioid alternative for targeted, on-demand pain relief in current hospital settings. A wearable cooling device with a more compact integration of its components helps achieve point-of-care pain management. By leveraging the latest advances in flexible and stretchable bioelectronics, an all-in-one nerve interface with cooling, temperature monitoring, and electrophysiological recordings may be possible in the future, offering real-time adjustment of cooling temperatures, duration, and intermittency to achieve desired physiological and therapeutic effects. ■

The flexible and bioresorbable microfluidic cooler for nerve blocks

With embedded microfluidics and microelectronics, this bioresorbable device can achieve precise and reversible pain mitigation through local coolant evaporation in the microfluidic channel, where the neighboring temperature sensor monitors real-time temperature changes.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. D. S. Goldberg, S. J. McGee, *BMC Public Health* **11**, 770 (2011).
2. N. B. Finnerup, *N. Engl. J. Med*. **380**, 2440 (2019).
3. M. Johnson, *Rev. Pain* **1**, 7 (2007).
4. B. M. Berman, H. M. Langevin, C. M. Witt, R. Dubner, *N. Engl. J. Med*. **363**, 454 (2010).
5. E. Ernst, V. Fialka, *J. Pain Symptom Manage*. **9**, 56 (1994).
6. J. T. Reeder *et al*., *Science* **377**, 109 (2022).
7. R. Janssen, *Neurosci. Biobehav. Rev*. **16**, 399 (1992).
8. T. Morgan *et al*., *J. Neurophysiol*. **123**, 2173 (2020).
9. S. G. Lomber, B. R. Payne, J. A. Horel, *J. Neurosci. Methods* **86**, 179 (1999).
10. E. Song, J. Li, S. M. Won, W. Bai, J. A. Rogers, *Nat. Mater*. **19**, 590 (2020).
11. M. M. Nahar *et al*., *Appl. Therm. Eng*. **194**, 117109 (2021).
12. A. Bar-Cohen, G. Sherwood, M. Hodes, G. Solbreken, *Microelectron. Reliab*. **37**, 502 (1997).
13. J. S. Chickos, W. E. Acree Jr., *J. Phys. Chem. Ref. Data* **32**, 519 (2003).
14. C. R. Ponce, S. G. Lomber, R. T. Born, *Nat. Neurosci*. **11**, 216 (2008).

10.1126/science.abm8159

TSUNAMIS

The wave blown around the world

The atmospheric wave from the Tonga eruption drove faster-than-normal tsunamis

By **Emily E. Brodsky** *and* **Thorne Lay**

On 15 January 2022, unusual waves appeared in Earth's atmosphere and oceans (*1*–*3*). The origin of the waves was clearly the catastrophic volcanic eruption in Tonga, which pummeled the atmosphere with the largest eruptive plume since the 1883 eruption of Krakatoa, Indonesia. On page 95 of this issue, Matoza *et al.* (*4*) show that the 2022 Tonga eruption generated waves in the water, air, and even in the ionosphere that wrapped around Earth multiple times. Tsunamis appeared to hop across the land into all of the major ocean basins. And on page 91 of this issue, Kubota *et al.* (*5*) explain that the tsunamis arrived much earlier than expected on the basis of conventional tsunami modeling, and the wave trains lasted much longer than for even the largest earthquakes (*5*).

Tsunamis are ocean waves with several-hundred-kilometer-long wavelengths that are usually caused by sudden perturbations of the ocean. Giant earthquakes, such as the 2004 magnitude 9.2 Sumatra earthquake, can produce abrupt seafloor movements that extend as much as 1500 km along the fault rupture on a tectonic plate boundary, displacing the overlying ocean and generating a devastating tsunami. But earthquakes are not the only events that trigger tsunamis—landslides and volcanic eruptions can also displace large volumes of water in the ocean and produce tsunami waves.

Once the water is displaced, tsunamis travel as ordinary gravity waves with buoyancy providing the restoring force. Gravity waves in the ocean include the familiar waves on the sea surface that break on the seashore. Tsunamis are usually governed by the same physics. However, because of their long wavelengths, the speed and path of the gravity waves are determined by the variable depth of the ocean. Investments in tsunami modeling, motivated by the events of the 2004 Indian Ocean tsunami and the 2011 Tohoku tsunami in Japan, have helped develop forecasting and warning systems for tsunamis that now produce reasonably accurate predictions after earthquakes of when and where tsunamis will occur as well as how high the waves will be.

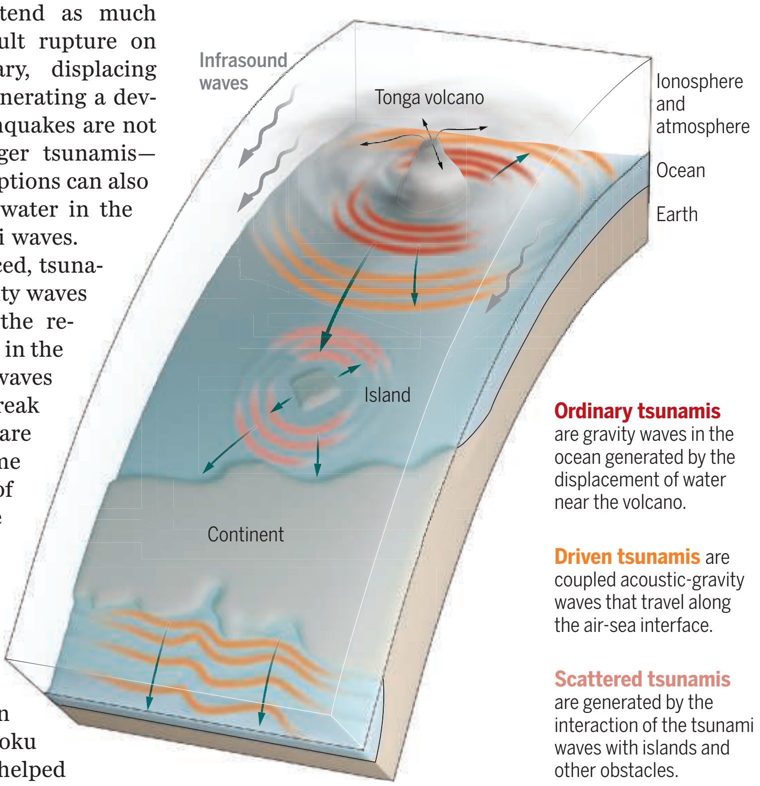
Surprisingly, the first tsunami arrivals produced by the Tonga eruption were observed several hours earlier than anticipated. Kubota *et al.* explain that the fast and complex tsunamis result from how the ocean interacts with the atmosphere after a volcanic eruption (see the figure). They show that the atmospheric and oceanic dynamics are intimately connected. The atmospheric waves were accompanied by a coupled oceanic wave that traveled at a speed that exceeded that of conventional tsunami gravity waves in the ocean (~200 m/s). The coupled wave can be modeled as an acoustic-gravity wave, which travels by simultaneously compressing (like a sound wave) while having an accompanying buoyancy force at the air-sea interface. In effect, the air wave blew the tsunami wave across the ocean, even passing over continents and driving new tsunamis when the ocean was again encountered.

This rare method of tsunami generation has a precedent in the last eruption that produced such strong atmospheric waves. Like the 2022 eruption in Tonga, the 1883 Krakatoa event also produced an intense atmospheric acoustic-gravity wave and infrasound that could be heard up to 3000 miles away (*4*). In 1955, Press and Ewing (*6*) noted that air-pressure increases were coincident with initial oceanic waves observed around the world. In 1967, Harkrider and Press (*7*) went on to analyze the coupled motion of the ocean and atmosphere to quantify the mechanistic connection. Of course, the data from 1883 were difficult to analyze and the Krakatoa observation stood as an oddity in the annals of tsunami generation until the 2022 eruption in Tonga produced extensive global observations (*4*).

The wavefield that resulted was captured with unprecedented resolution as explained by Matoza *et al.* The rich data make it clear that very large-scale atmospheric transients of eruptions or atmospheric impacts should be expected to produce large, fast, and complex tsunamis around the world. The volcanic blast produced a nondispersive wave in the atmosphere that traveled along Earth's surface with a pressure oscillation lasting over 2000 s that expanded away from the source at about 315 m/s (*4*, *5*, *8*–*11*). Slower acoustic waves were of high enough frequencies (>20 Hz infrasound) that they could be heard by people in Alaska, more than 6000

Multiple tsunami generation mechanisms

The January 2022 Tonga eruption generated waves in the oceans, atmosphere, and ionosphere as reported by Matoza *et al.*, including unusual tsunamis blown by atmosphere at the air-sea interface as shown by Kubota *et al*.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Department of Earth and Planetary Sciences, University of California Santa Cruz, CA, USA Email: brodsky@ucsc.edu

miles away from the erupting volcano.

In addition, conventional tsunami waves were generated near the source in response to the initial ocean displacement, and scattered tsunami waves were generated as the driven wave encountered variations in ocean depth across the Pacific. This combination of generation mechanisms produced the very long-lasting sequence of tsunami waves, including the largest waves, which were twice as large as the initial atmospherically coupled one (*11*). Even more elastic waves were produced in the solid earth and were observed on seismometers globally (*12*).

The 2022 Tonga eruption reawakens curiosity about processes crossing traditional boundaries between Earth systems. Atmosphere–Ocean–Solid Earth coupling is a major focus of climate research, yet observations that directly sample their interaction may have been underutilized. Volcanic eruption studies show that the tools of infrasound and seismology that have traditionally been used to study the solid Earth potentially have much broader uses. New technologies and networks in a data-rich world are providing windows into dynamic interactions between Earth's systems. Oceans, rivers, glaciers, and hurricanes all produce distinctive waves that are now routinely recorded on global networks (*13*–*15*). Earth's systems are broadcasting their interactions through multiple types of waves, and now instrumentation is in place to receive their signals. It is up to the scientific community to reach out across the disciplinary boundaries and use the data to ask the right questions about the entire Earth system. ■

REFERENCES AND NOTES

1. D. Adam, *Nature* **601**, 497 (2022).
2. D. R. Themens *et al.*, *Geophys. Res. Lett.* **49**, 2022GL098158 (2022).
3. D. A. Yuen *et al.*, *Earthquake Res. Adv.* 10.1016/j.eqrea.2022.100134 (2022).
4. R. S. Matoza *et al.*, *Science* **377**, 95 (2022).
5. T. Kubota *et al.*, *Science* **377**, 91 (2022).
6. M. Ewing, F. Press, *Trans. Am. Geophys. Union* **36**, 53 (1955).
7. D. Harkrider, F. Press, *Geophys. J. Int.* **13**, 149 (1967).
8. S. Otsuka, *Geophys. Res. Lett.* **49**, e2022GL098324 (2022).
9. A. Amores *et al.*, *Geophys. Res. Lett.* **49**, e2022GL098240 (2022).
10. J.-T. Lin *et al.*, *Geophys. Res. Lett.* **49**, e2022GL098222 (2022).
11. M. Carvajal, I. Sepúlveda, A. Gubler, R. Garreaud, *Geophys. Res. Lett.* **49**, 2022GL098153 (2022).
12. P. Poli, N. M. Shapiro, *Geophys. Res. Lett.* **49**, e2022GL098123 (2022).
13. K. L. Cook, M. Dietze, *Annu. Rev. Earth Planet. Sci.* **50**, 183 (2022).
14. H. R. Shaddox, E. E. Brodsky, S. R. Ramp, K. A. Davis, *AGU Adv.* **2**, e2021AV000475 (2021).
15. T. Tanimoto, A. Lamontagne, *Geophys. Res. Lett.* **41**, 7532 (2014).

10.1126/science. abq5392

NEURODEGENERATION

Elucidating the causes of neurodegeneration

Investigating phase separation in neurodegeneration highlights evidence needed for causation

By **Gregory A. Petsko**[1] *and* **Scott A. Small**[2]

How can a given gene, or gene product, or the cellular pathway it serves be determined as causal for neurodegenerative diseases instead of being a consequence of the underlying pathology? There are questions about the current theories that explain neurodegeneration, with therapeutic strategies producing disappointing results. On page 46 of this issue, Zhou *et al.* (*1*) attempt to address whether the aberrant phase separation of protein domains of low amino acid sequence complexity are causally pathogenic in several neurodegenerative diseases, including a form of frontotemporal dementia. Their approach has potential application in identifying causative factors in a range of disorders, as well as to the formation of membraneless organelles in normal cells.

In 1884, Robert Koch and Friedrich Loeffler proposed postulates to establish that a microorganism caused an infectious disease (*2*). Despite many iterations of Kochs' postulates since (*3*), the concept of a set of criteria that should be satisfied to determine causality for an infectious disease remains useful. Nowhere is such rigor more needed than in the case of neurodegenerative diseases. The complexity, chronicity, and relative inaccessibility to experimentation of the human brain, and lack of good animal models, as well as the phenotypic heterogeneity and age dependence of most of these disorders make finding causal biological pathways very difficult. Consequently, specific criteria to establish causality for a neurodegenerative disease and to validate a target and therapeutic approach might be useful. Ideally, given the complexity of the problem, such postulates should reflect an integrative biology approach (see the figure).

A set of criteria are proposed, integrating evidence from human molecular biology (whether key molecules of the pathway are dysregulated in the disease), cellular biology (when disrupted, whether the pathway causes the disease's cardinal cellular pathologies), and anatomical biology (when disrupted, whether the pathway causes the disease's cardinal anatomical pathology). In the case of Alzheimer's disease, the hypothesis that misprocessing of the amyloid-β precursor protein (APP) produces cytotoxic extracellular amyloid plaques fulfills parts of the first two criteria. But amyloid plaque deposition is not primarily observed in the trans-entorhinal cortex (TEC), where neurodegeneration begins in Alzheimer's disease—an anatomical mismatch between plaques and neurodegeneration noted by Alzheimer himself (*4*). The TEC is important in short-term memory and navigation, which become disrupted in Alzheimer's disease. Two promising new hypotheses for the causes of neurodegeneration in Alzheimer's disease—dysfunction of the innate immune system in the brain (*5*) and defects in endosomal protein trafficking (*6*)—are not mutually exclusive; the latter may fulfill all three criteria (*7*).

Zhou *et al.* describe experiments aimed at satisfying these criteria in an attempt to resolve an ongoing controversy about how low-complexity sequence domains form membraneless organelles (*8*) but also to identify a causal mechanism for several neurodegenerative diseases. It has been known for more than a century that eukaryotic cells contain phase-separated liquid-like droplets, but interest in their composition and function waned. However, in 2009, germline P granules in *Caenorhabditis elegans* were shown to be liquid droplets that displayed similar behavior to that of oil in water (*9*). Further work revealed that such phase separation was common, especially for some protein-RNA complexes, and it was proposed that the resulting membraneless organelles likely played important functional roles in the regulation of gene expression and nuclear import (*10*) by providing a way to temporarily organize critical components without the complexity of transporting molecules across a membrane.

In 2012, amyloid-like fibers were identified in hydrogels formed from low-complex-

[1]Ann Romney Center for Neurologic Diseases, Department of Neurology, Brigham & Women's Hospital and Harvard Medical School, Boson, MA, USA. [2]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, NY, USA. Email: gpetsko@bwh.harvard.edu; SAS68@columbia.edu

miles away from the erupting volcano.

In addition, conventional tsunami waves were generated near the source in response to the initial ocean displacement, and scattered tsunami waves were generated as the driven wave encountered variations in ocean depth across the Pacific. This combination of generation mechanisms produced the very long-lasting sequence of tsunami waves, including the largest waves, which were twice as large as the initial atmospherically coupled one (*11*). Even more elastic waves were produced in the solid earth and were observed on seismometers globally (*12*).

The 2022 Tonga eruption reawakens curiosity about processes crossing traditional boundaries between Earth systems. Atmosphere–Ocean–Solid Earth coupling is a major focus of climate research, yet observations that directly sample their interaction may have been underutilized. Volcanic eruption studies show that the tools of infrasound and seismology that have traditionally been used to study the solid Earth potentially have much broader uses. New technologies and networks in a data-rich world are providing windows into dynamic interactions between Earth's systems. Oceans, rivers, glaciers, and hurricanes all produce distinctive waves that are now routinely recorded on global networks (*13*–*15*). Earth's systems are broadcasting their interactions through multiple types of waves, and now instrumentation is in place to receive their signals. It is up to the scientific community to reach out across the disciplinary boundaries and use the data to ask the right questions about the entire Earth system. ■

REFERENCES AND NOTES

1. D. Adam, *Nature* **601**, 497 (2022).
2. D. R. Themens *et al.*, *Geophys. Res. Lett*. **49**, 2022GL098158 (2022).
3. D. A. Yuen *et al.*, *Earthquake Res. Adv*. 10.1016/j.eqrea.2022.100134 (2022).
4. R. S. Matoza *et al.*, *Science* **377**, 95 (2022).
5. T. Kubota *et al.*, *Science* **377**, 91 (2022).
6. M. Ewing, F. Press, *Trans. Am. Geophys. Union* **36**, 53 (1955).
7. D. Harkrider, F. Press, *Geophys. J. Int*. **13**, 149 (1967).
8. S. Otsuka, *Geophys. Res. Lett.* **49**, e2022GL098324 (2022).
9. A. Amores *et al.*, *Geophys. Res. Lett*. **49**, e2022GL098240 (2022).
10. J.-T. Lin *et al.*, *Geophys. Res. Lett*. **49**, e2022GL098222 (2022).
11. M. Carvajal, I. Sepúlveda, A. Gubler, R. Garreaud, *Geophys. Res. Lett.* **49**, 2022GL098153 (2022).
12. P. Poli, N. M. Shapiro, *Geophys. Res. Lett.* **49**, e2022GL098123 (2022).
13. K. L. Cook, M. Dietze, *Annu. Rev. Earth Planet. Sci.* **50**, 183 (2022).
14. H. R. Shaddox, E. E. Brodsky, S. R. Ramp, K. A. Davis, *AGU Adv.* **2**, e2021AV000475 (2021).
15. T. Tanimoto, A. Lamontagne, *Geophys. Res. Lett*. **41**, 7532 (2014).

10.1126/science.abq5392

NEURODEGENERATION

Elucidating the causes of neurodegeneration

Investigating phase separation in neurodegeneration highlights evidence needed for causation

By **Gregory A. Petsko**[1] *and* **Scott A. Small**[2]

How can a given gene, or gene product, or the cellular pathway it serves be determined as causal for neurodegenerative diseases instead of being a consequence of the underlying pathology? There are questions about the current theories that explain neurodegeneration, with therapeutic strategies producing disappointing results. On page 46 of this issue, Zhou *et al.* (*1*) attempt to address whether the aberrant phase separation of protein domains of low amino acid sequence complexity are causally pathogenic in several neurodegenerative diseases, including a form of frontotemporal dementia. Their approach has potential application in identifying causative factors in a range of disorders, as well as to the formation of membraneless organelles in normal cells.

In 1884, Robert Koch and Friedrich Loeffler proposed postulates to establish that a microorganism caused an infectious disease (*2*). Despite many iterations of Kochs' postulates since (*3*), the concept of a set of criteria that should be satisfied to determine causality for an infectious disease remains useful. Nowhere is such rigor more needed than in the case of neurodegenerative diseases. The complexity, chronicity, and relative inaccessibility to experimentation of the human brain, and lack of good animal models, as well as the phenotypic heterogeneity and age dependence of most of these disorders make finding causal biological pathways very difficult. Consequently, specific criteria to establish causality for a neurodegenerative disease and to validate a target and therapeutic approach might be useful. Ideally, given the complexity of the problem, such postulates should reflect an integrative biology approach (see the figure).

A set of criteria are proposed, integrating evidence from human molecular biology (whether key molecules of the pathway are dysregulated in the disease), cellular biology (when disrupted, whether the pathway causes the disease's cardinal cellular pathologies), and anatomical biology (when disrupted, whether the pathway causes the disease's cardinal anatomical pathology). In the case of Alzheimer's disease, the hypothesis that misprocessing of the amyloid-β precursor protein (APP) produces cytotoxic extracellular amyloid plaques fulfills parts of the first two criteria. But amyloid plaque deposition is not primarily observed in the trans-entorhinal cortex (TEC), where neurodegeneration begins in Alzheimer's disease—an anatomical mismatch between plaques and neurodegeneration noted by Alzheimer himself (*4*). The TEC is important in short-term memory and navigation, which become disrupted in Alzheimer's disease. Two promising new hypotheses for the causes of neurodegeneration in Alzheimer's disease—dysfunction of the innate immune system in the brain (*5*) and defects in endosomal protein trafficking (*6*)—are not mutually exclusive; the latter may fulfill all three criteria (*7*).

Zhou *et al.* describe experiments aimed at satisfying these criteria in an attempt to resolve an ongoing controversy about how low-complexity sequence domains form membraneless organelles (*8*) but also to identify a causal mechanism for several neurodegenerative diseases. It has been known for more than a century that eukaryotic cells contain phase-separated liquid-like droplets, but interest in their composition and function waned. However, in 2009, germline P granules in *Caenorhabditis elegans* were shown to be liquid droplets that displayed similar behavior to that of oil in water (*9*). Further work revealed that such phase separation was common, especially for some protein-RNA complexes, and it was proposed that the resulting membraneless organelles likely played important functional roles in the regulation of gene expression and nuclear import (*10*) by providing a way to temporarily organize critical components without the complexity of transporting molecules across a membrane.

In 2012, amyloid-like fibers were identified in hydrogels formed from low-complex-

[1]Ann Romney Center for Neurologic Diseases, Department of Neurology, Brigham & Women's Hospital and Harvard Medical School, Boson, MA, USA. [2]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, NY, USA. Email: gpetsko@bwh.harvard.edu; SAS68@columbia.edu

ity domains (*11*). Low-complexity domains are stretches of protein sequence comprising only a few different amino acids; they have been found in most of the proteins in membraneless organelles. Many have also been linked to neurologic disorders, the prevailing hypothesis being that this self-association sequesters critical proteins from their normal function and recruits other key cellular components into the fibril. But unlike the amyloid fibrils observed in Alzheimer's disease and other neurodegenerative diseases, the fibers in hydrogels had few interfacial hydrophobic residues and were highly labile.

Controversy has arisen around the role of secondary and tertiary structures in the self-association of low-complexity domains and consequent phase separation. Two competing models have been proposed. In one (*12*), low-complexity domain self-association is driven by the formation of a labile cross–β sheet structure stabilized by hydrogen bonds between N–H and C=O groups of adjacent strands of the polypeptide backbone. Because the polypeptide backbone is the same for each residue, in this model specificity as to which chains associate, and in what register, arises from some side chain interactions, as observed in crystal structures of amyloid-like fibers (*13*). In the other model (*14*), specific secondary structure plays little or no role, and backbone interactions are not thought to be essential; coalescence is proposed to be driven by stabilizing interactions between specifically spaced side chains.

Zhou *et al.* have, in essence, applied the molecular biology and cell biology criteria to demonstrate that formation of the cross–β sheet structure is the mechanism that drives both membraneless organelle formation and pathogenesis in certain neurodegenerative diseases. In so doing, they solve a thorny problem: how to test the role of the polypeptide chain backbone when virtually any mutation will leave that backbone intact. They stitched synthetic peptides with one modified backbone residue into the middle of the low-complexity domain in the RNA binding protein TDP-43 (TAR DNA-binding protein 43). TDP-43 mutations cause amyotropic lateral sclerosis and frontotemporal dementia. The authors made 24 synthetic peptides, each of which put a methyl cap on a single-peptide backbone nitrogen atom, blocking its ability to form the interstrand hydrogen bonds of a β sheet. By adding this modification across the predicted cross–β sheet–forming region, they map the precise region that requires backbone hydrogen bonding and confirm that the labile cross–β sheet structure mediates self-association. That removal of a single backbone hydrogen bond could disrupt the entire structure shows just how weak—and therefore dynamic and reversible—the self-association actually is.

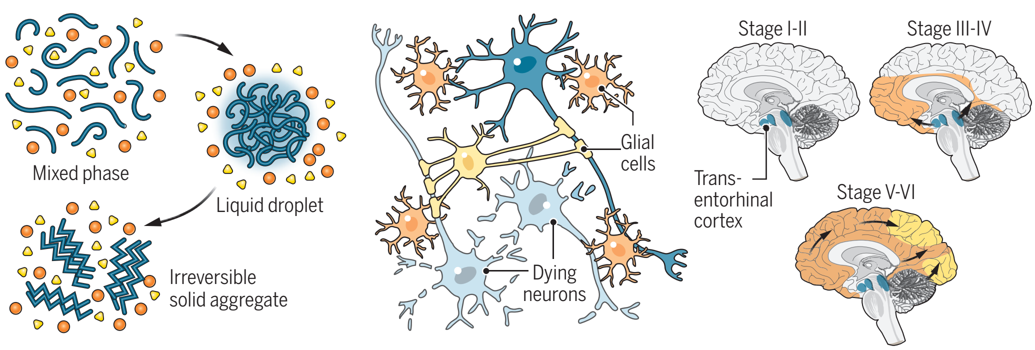
The authors go on to resolve how various low-complexity domain missense mutations in three different proteins are causative of several neurodegenerative diseases, showing that all of them add stability to their cognate structures in a manner understandable through the lens of structural biology. Noting that mutating Pro^{320} in the TDP-43 low-complexity domain to Gly, Ser, and Ala (proline has no free backbone N–H to form an interstrand hydrogen bond; all other residues do) caused precipitation and stable fiber formation, Zhou *et al.* looked for disease-causing mutations in proline residues in other low-complexity domains (the molecular biology criterion). They found such mutations in neurofilament light chain (NFL): When Pro^{22} or Pro^{8} are mutated, NFL forms stable aggregates that cause Charcot-Marie-Tooth disease (an inherited peripheral nerve disorder). Moreover, capping the residues replacing the proline with backbone-nitrogen methyl groups restored normal NFL behavior in biochemical assays.

Similar analyses of disease-causing mutations in prolines in the tau protein, which give rise to frontotemporal dementia, and in heterogeneous nuclear ribonucleoprotein A2 (hnRNPA2), which cause Paget's disease, are consistent with the model that cross–β sheet formation is important for low-complexity domain labile self-association. The results also suggest that making these dynamic structures too stable—for example, when the strategically placed proline residues that keep the structure labile are mutated—is pathogenic because the dynamic assembly and disassembly of low-complexity domain proteins with their partners is essential for the normal function of the cell. Work such as that of Zhou *et al.*, as well as by others, on protein aggregation has brought new rigor to understanding neurodegenerative diseases, and consequently, prospects for finding definitive causes and effective therapies are brighter. At least for some disorders, finding ways to maintain reversible phase separation of low-complexity domains now seems an attractive strategy. ■

Integrative biology as evidence for causation in neurodegeneration

Three main criteria should be met for establishing causation in neurodegenerative diseases: Key molecules of a biological pathway should be dysregulated in the disease (molecular biology), and when disrupted, the pathway should cause the disease's cardinal cellular pathologies (cellular biology) and explain the disease's anatomical vulnerability (anatomical biology).

Molecular biology | **Cellular biology** | **Anatomical biology**



For example, mutated low-complexity domain proteins form less dynamic membraneless organelles in several neurodegenerative diseases.

Neurodegeneration and neuroinflammation are cardinal cellular pathologies in neurodegenerative diseases.

For example, neurodegeneration begins in the trans-entorhinal cortex in Alzheimer's disease and then spreads across the brain.

REFERENCES AND NOTES

1. X. Zhou *et al.*, *Science* **377**, eabn5582 (2022).
2. R. Koch, *Verhandlungen des X Internationalen Medichinischen Congresses, Berlin* **1**, 35 (1890).
3. A. L. Byrd, J. A. Segre, *Science* **351**, 224 (2016).
4. A. Alzheimer, *Z. Gesamte Neurol. Psychiatr.* **4**, 356 (1911).
5. Y. Shi, D. M. Holtzman, *Nat. Rev. Immunol.* **18**, 759 (2018).
6. S. A. Small, G. A. Petsko, *Sci. Transl. Med.* **13**, eabb1717 (2020).
7. S. Simoes *et al.*, *Cell Rep.* **37**, 110182 (2021).
8. M. Kato, X. Zhou, S. L. McKnight, *RNA* **28**, 3 (2022).
9. C. P. Brangwynne *et al.*, *Science* **324**, 1729 (2009).
10. S. Frey, D. Görlich, *Cell* **130**, 512 (2007).
11. M. Kato *et al.*, *Cell* **149**, 753 (2012).
12. S. Xiang *et al.*, *Cell* **163**, 829 (2015).
13. E. L. Guenther *et al.*, *Nat. Struct. Mol. Biol.* **25**, 463 (2018).
14. K. A. Burle *et al.*, *Mol. Cell* **50**, 231 (2015).

ACKNOWLEDGMENTS

G.A.P. and S.A.S. are cofounders of Retromer Therapeutics, which is developing treatments for neurodegenerative diseases.

10.1126/science.adc9969

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

VIEWPOINT: COVID-19

The hiatus of the handshake

Human contact has been altered in ways that may affect endemic infections for years to come

By **Kristin Nelson**[1] *and* **Ben Lopman**[1,2]

The response to the COVID-19 pandemic prompted abrupt and potentially lasting changes to human behavior, including the types of direct contact that enable transmission of common pathogens. Although the handshake and other types of physical contact are gradually returning, they have been on an extended hiatus. Such changes have already altered the epidemiology of a broad range of infectious diseases, including influenza, measles, and norovirus, and will likely continue to affect their age distribution, severity, and typical seasonal patterns. Changes to contact patterns may also nudge the evolutionary trajectory of pathogens, as they adapt to new norms of less human contact. Initiatives to measure social contact patterns through time and track their effects on the epidemiology of endemic pathogens are essential to both manage the potential for resurgence of common infections and to reenergize control efforts.

There are two fundamental determinants of the dynamics of directly transmitted infectious diseases: the human contact that enables transmission and a pathogen's biological natural history. Reductions in human contact that occurred during the COVID-19 pandemic have substantially affected patterns of circulation of other pathogens (see the figure). Reported incidence of most directly transmitted infections fell to virtually zero in early 2020 amid initial lockdown measures. Although these declines may have partially been due to gaps in reporting, they mostly reflect true reductions in disease incidence. As restrictions were relaxed, circulation of certain infections resumed—in some cases, with shifted seasonal patterns—but they have yet to return to their pre-pandemic levels.

Despite the widespread relaxation of nonpharmaceutical interventions, a return to pre-pandemic "normal" has yet to fully materialize. Studies of specific populations in the US are consistent with those from general European populations, which report contacts averaging less than five per day in early 2022—less than half of pre-pandemic levels (*1*–*3*). This reduction so far seems sustained, despite a return to workplaces and the resumption of in-person schooling, social activities, and large gatherings.

Not only are people interacting less, but where those interactions take place has shifted. The move toward increased remote work options has de-densified workplaces, reducing the number of contacts reported at work, and is likely to reduce the contribution of these settings to disease transmission. By contrast, school-based contacts are likely to return to typical, pre-pandemic patterns (*4*). When they occur, interactions may be somewhat less likely to transmit infection than those that occurred prior to the pandemic. Data show that contacts are less likely to involve physical touch now than before the pandemic and that masking remains common: ~30% of the US population reports routinely masking at work, and ~25% of children are still masking at school (*5*). However, studies from other countries suggest that these measures are now less common, underscoring the need to better understand how behaviors differ by location (*6*). Although it is uncertain how long current levels of masking and physical distancing will continue, the widespread normalization of this behavior suggests that they are likely to continue in some form, perhaps on public transport, in health care settings, and among people at high risk for severe respiratory illness.

Even if these reductions in contact are modest, impacts on endemic diseases may be substantial. Many pathogens maintain a fine balance of endemic equilibrium; that is, with an effective reproduction number (the number of cases caused by an infection in a population with some immunity) close to 1. Lower contact rates mean reduced transmission and a disruption of this balance for the many pathogens that are transmitted directly from person to person. A near-complete return to pre-pandemic contact rates after 2 years could result in considerable changes to transmission dynamics. Some pathogens that confer lifelong immunity may resurge to higher than pre-pandemic levels after an extended period of low circulation, resulting from a buildup of susceptible individuals, before settling into an endemic state at lower incidence. Other infections may display several years of altered seasonal patterns. Indeed, this is consistent with out-of-season respiratory syncytial virus (RSV) epidemics already documented in summer 2021. For other infections, however, it is likely that the current period of very low incidence may be followed by resurgence, and historical data may provide insight into what to expect.

For childhood infections that ceased to circulate among the youngest children over the past several years, first infections will occur at older ages and disease profiles for some may be less severe. For example, the generally lower severity of dehydrating diarrhea in a 5-year-old compared with an infant under 1 year of age. But for some infections, such as varicella (chickenpox), disease is more likely to be severe at older ages. The increasing severity of poliomyelitis in the US in the 1950s was most likely the result of improvements in sanitation that reduced environmental transmission and the risk of exposure at younger ages, which drove infections into older age groups in which paralysis is more common (*7*).

For infections with strong seasonal patterns, persistently lower contact rates may shift the timing of peaks. For example, lower levels of transmission may result in a delayed peak of seasonal infections, resulting in stresses on the health care system at new times of the year. Any intervention that reduces circulation of a pathogen may result in changes to its dynamics. For example, the introduction of rotavirus vaccination—which protected vaccinated children and reduced transmission—in 2006 reduced overall rotavirus incidence and has since perturbed seasonal patterns (*8*).

Sustained changes in human contact may also have the potential to drive the evolution of endemic pathogens. Those with high transmissibility, either because of a lower average infectious dose or a better survival profile outside the host, could have an increasingly strong evolutionary advantage. Alternatively, pathogens with more prolonged infectious periods that lead to more opportunities for transmission may also be more successful in a new normal defined by less human-to-human contact. Changes in the severity of endemic infections due to evolutionary pressures may be harder to anticipate. The COVID-19 pandemic underscores that the evolutionary trajectory of pathogens does not always follow predictable patterns. The emergence of the Delta severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant, which was more severe than its precursor, Alpha, was closely followed by less-severe Omicron variants (*9*, *10*), illustrating that

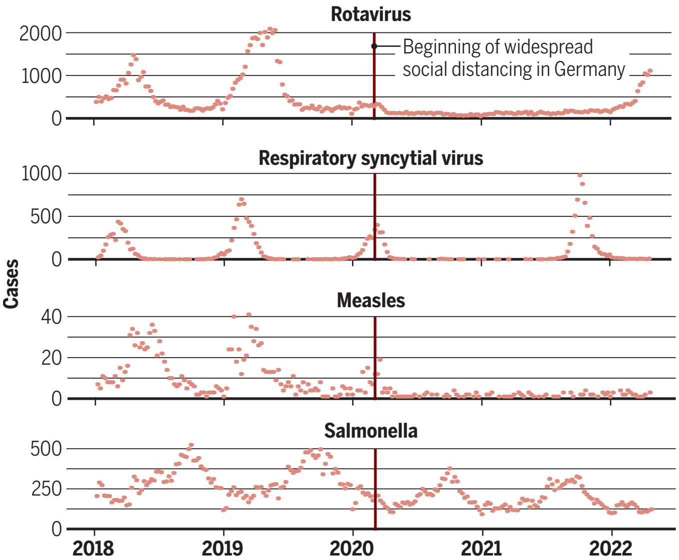
[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [2]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. Email: blopman@emory.edu

the evolution of virulence is not constrained in one direction.

For pathogens that employ multiple routes of transmission, variants that are transmitted via environmental routes, such as contaminated water, soil, or objects, might outcompete those that rely on direct contact. For example, genogroup 1 noroviruses, which are more often transmitted by food and water, may increase in prevalence relative to genogroup 2 noroviruses, which are predominantly transmitted through person-to-person contact (*11*). Notably, the incidence of enteric pathogens such as *Salmonella,* which transmit through routes that do not require direct person-to-person contact, were less affected by social distancing. In the most extreme scenario, fewer infections may slow or even stop the evolution of endemic infections if the effective reproductive number is suppressed below 1 for a sufficient amount of time. During the lull in influenza virus circulation in the wake of the pandemic, the influenza B/Yamagata lineage has virtually disappeared, leading to speculation that it has gone extinct (*12*). The current period of low incidence may also present a rare opportunity to eliminate local transmission of eradicable pathogens, such as measles. In these cases, redoubled vaccination efforts are also required to overcome immunity gaps and temporary falls in vaccine coverage that occurred during the pandemic.

The effects of reduced contact on diseases caused by endemic pathogens

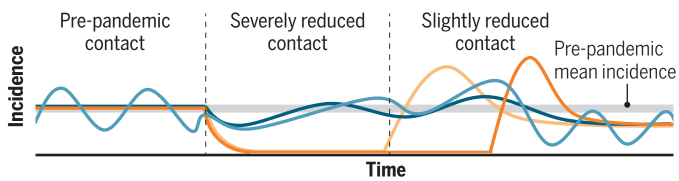
Weekly case reports of a selection of common pathogens in Germany are shown from January 2018 through May 2022 [source: SurvStat database (https://survstat.rki.de/)].



Possible future trajectories

Accrual of susceptible people from reduced contact may cause resurgence, which is slower for illnesses against which populations have vaccine-induced protection. Pathogens for which the susceptible population is maintained by loss of immunity are less affected. Seasonal infections may display years of disrupted patterns.

Lifelong immunity
...with vaccination
Temporary immunity
...and seasonal



To ascertain whether changes in social contact will be long-standing and broadly impactful on infectious disease epidemiology, two data streams are pertinent. The first is direct measurement of contact. Despite human contact being the proximate social determinant of infectious disease transmission, it is rarely measured in a way that ensures population representativeness, and even more rarely, longitudinally. In Europe, the CoMix study has provided near–real time data during the pandemic through large, weekly online surveys on the number and nature of contacts (*2*), but no such data source exists for the US or many other countries. Notably, there is little to no data available from low- and middle-income countries, where the burden of many of these infections is highest, representing a key data gap. Instead, data describing contact patterns in specific groups (*4*, *13*, *14*) and proxies for social contact, including mobility data that is passively collected from mobile devices, has been heavily relied on. Geolocation data from mobile devices, which can be used to infer the proximity among people, may provide a more accurate accounting of social contact that is relevant for transmission risk (*15*). However, such data lack information on the physicality of contact and whether masks were worn.

The second data stream required to understand how changing social contact affects disease epidemiology is case surveillance data, providing a direct measure of disease patterns and trends in the sizes and locations of outbreaks. A key metric to be monitored is the average age and severity of cases of a given infectious disease. Lower transmission, because of fewer potentially infectious contacts, will result in an older average age of cases. Of course, careful interpretation is critical because multiple factors are at play—key among them, changing health care seeking and testing practices, which can bias traditional surveillance data. Nonetheless, combining robust case surveillance with genomic data and phylogenetic analysis can provide signals of pathogen evolution and their potential impact on public health.

At its core, infectious disease epidemiology is the study of person, place, and time: who becomes sick, where infection and transmission occur, and when disease incidence rises or falls. Substantial changes to human behavior in response to the COVID-19 pandemic stand to affect all three, producing ripple effects for many endemic diseases and upending the current understanding of their epidemiology. To adapt, modern surveillance activities should do more than count cases, it should include human behavioral and pathogen genomic data to provide a rich picture of the changing epidemiology of endemic pathogens. The scientific community has spent the past 2 years characterizing the epidemiology of a novel infection—it is time to ensure that disease surveillance systems will be ahead of more familiar foes. ■

REFERENCES AND NOTES

1. K. Prem *et al*., *PLOS Comput. Biol*. **13**, e1005697 (2017).
2. A. Gimma *et al*., *PLOS Med*. **19**, e1003907 (2022).
3. C. Y. Liu *et al*., *Epidemiology* **32**, 781 (2021).
4. M. C. Kiti *et al*., *Epidemics* **36**, 100481 (2021).
5. Kaiser Family Foundation, KFF COVID-19 Vaccine Monitor; http://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/.
6. D. Thindwa *et al*., medRxiv 10.1101/2021.12.16.21267959 (2021).
7. N. Nathanson, J. R. Martin, *Am. J. Epidemiol*. **110**, 672 (1979).
8. J. E. Tate *et al*., *Pediatr. Infect. Dis. J*. **32**, 741 (2013).
9. J. A. Lewnard *et al*., *Nat. Med*. 10.1038/s41591-022-01887-z (2022).
10. K. A. Twohig *et al*., *Lancet Infect. Dis*. **22**, 35 (2022).
11. L. Verhoef *et al*., *Emerg. Infect. Dis*. **21**, 592 (2015).
12. M. Koutsakos *et al*., *Nat. Rev. Microbiol*. **19**, 741 (2021).
13. T. Chin *et al*., medRxiv 10.1101/2021.09.25.21264082 (2021).
14. D. M. Feehan, A. S. Mahmud, *Nat. Commun*. **12**, 893 (2021).
15. O. Morozova *et al*., *Sci. Rep*. **11**, 20271 (2021).

ACKNOWLEDGMENTS

The authors thank J. de Roode for helpful comments. B.L. serves as a consultant to Epidemiological Research and Methods, LLC.

10.1126/science.abp9316

GRAPHIC: N. DESAI/*SCIENCE*

POLICY FORUM

DIVERSITY

Can rubrics combat gender bias in faculty hiring?

Some bias persisted, but rubric use should be encouraged

By **Mary Blair-Loy[1], Olga V. Mayorova[1], Pamela C. Cosman[2], Stephanie I. Fraley[3]**

Research has documented the presence of bias against women in hiring, including in academic science, technology, engineering, and mathematics (STEM). Hiring rubrics (also called criterion checklists, decision support tools, and evaluation tools) are widely recommended as a precise, cost-effective remedy to counteract hiring bias, despite a paucity of evidence that they actually work (see table S8). Our in-depth case study of rubric usage in faculty hiring in an academic engineering department in a very research-active university found that the rate of hiring women increased after the department deployed rubrics and used them to guide holistic discussions. Yet we also found evidence of substantial gender bias persisting in some rubric scoring categories and evaluators' written comments. We do not recommend abandoning rubrics. Instead, we recommend a strategic and sociologically astute use of rubrics as a department self-study tool within the context of a holistic evaluation of semifinalist candidates.

Although academic STEM aspires to be a meritocracy, its taken-for-granted cultural schemas of merit smuggle in biases (*1*), which contribute to a dearth of diversity that undermines scientific innovation and impact (see table S8). In academic engineering, one of the most male-dominated STEM fields, on average 17.6% of engineering faculty positions are held by women (*2*). Although the percent of women engineering doctorates increased from 15.8% in 2000 to 24% in 2019 (*3*), these increases will not be matched by gains in the professoriate if women face unfair barriers at hiring.

Academic policy-makers and EDI (equity, diversity, and inclusion) specialists strongly encourage rubric usage, in which faculty evaluators systematically rate each candidate on a set of previously agreed-on criteria. This process is believed to counteract the bias of individual evaluators by promoting slower, more deliberative, and analytical thinking and by focusing them on skill sets that directly affect job performance rather than on impressions and intuitions (*4*, *5*). However, we are aware of only one study, conducted in a laboratory setting, in which participants rated candidate summaries, which shows that agreeing on rubric criteria in advance reduces evaluation bias (*6*). We are aware of no studies that analyze the effect of rubric use on bias in real-world hiring, in which actual evaluators assess voluminous candidate files to make actual high-stakes decisions. Real-world case studies are important because the effectiveness of interventions depends on the social context and the identities of all involved (*7*).

Despite this paucity of evidence, many fields have developed rubrics to standardize candidate assessment and have promoted rubrics as a best practice for EDI in hiring. In policy guides for academic hiring, several applied treatises and websites provide sample rubrics (see table S8) (*5*). A recent influential review of faculty hiring lists "mandatory use" of rubrics as one of the interventions that the authors "view as having the most promise for [university] institutions seeking to improve inclusivity in hiring across disciplines" (*4*).

CASE STUDY OF RUBRIC USAGE

To help address this knowledge and policy gap, we developed an in-depth case study of an engineering department in a research-intensive (Carnegie classification R1), highly ranked university. Like other R1s, this department strongly values research productivity when evaluating faculty candidates (*8*). Like most academic engineering departments, our case department was male dominated; women composed 18% of the faculty, which is close to the 17.6% national average (*2*).

Rubric usage and hiring patterns

We started with a faculty candidate evaluation template from the University of Michigan STRIDE program (Strategies and Tactics for Recruiting to Improve Diversity and Excellence), which is funded by the National Science Foundation (NSF). This template includes widely accepted criteria for faculty at research-intensive universities (*5*).

We worked with the department under study to adapt the template to fit its searches. Its rubric evaluated faculty candidates across six dimensions: research productivity, research impact, teaching ability, contributions to diversity, potential for collaboration, and overall impression (see fig. S1). Rubric scores ranged from excellent to poor for each evaluation category (we translated these ratings into a numerical variable: excellent = 4, good = 3, neutral = 2, fair = 1, and poor = 0). Written commentary was also encouraged.

Department faculty agreed to fill out the rubric as a tool in their evaluation of the semifinalist list compiled by the recruitment committee. In four faculty search cycles over four recent academic years, faculty used the rubric to evaluate written materials supporting applications of 62 semifinalists (32 women and 30 men; gender was self-reported by candidates in their application file). At the beginning of each faculty meeting that was focused on selecting finalists, a faculty member summarized and presented rubric scoring results and commentary, with evaluators anonymized, filtering out any inaccurate or off-topic content.

Our analysis next compared the proportion of women hired during the 8-year period immediately before rubric use (which we refer to as "Phase One") with the proportion hired during the four academic years of rubric use ("Phase Two"). Near the outset of Phase One, the campus implemented three EDI interventions: Faculty serving as equity advisers took on administrative oversight of shortlisted candidates, applicants' "Contributions to Diversity Statements" (C2D) were added to application files, and equity advisers began giving diversity training to search committee members. Such training focused on evaluating C2D statements and an overview of research on implicit biases. Nonetheless, during Phase One, the department conducted eight searches and hired eight men and one woman.

At the outset of Phase Two, the department introduced an additional intervention: rubrics used as described above. During Phase Two, the department conducted four faculty searches and hired three women and six men. The number of women hired increased from only one per nine hires in Phase One to three per nine hires in Phase

[1]Department of Sociology, University of California, San Diego, La Jolla, CA, USA. [2]Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, USA. [3]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. Email: mblairloy@ucsd.edu; sifraley@ucsd.edu

Two. The phase with the increased hiring of women coincides with the period when rubrics were used. (We cannot fully attribute this increase to rubric utilization because unmeasured factors may also have changed.)

Analysis of rubric scoring

All 62 semifinalist candidates received rubric scores from 6 to 21 faculty, with a mean of 13.5 and a median of 12. There was no statistically significant gender difference in the number of scores received.

Analysis of the rubric scoring patterns for men and women candidates revealed statistically significant differences in three of the six evaluation categories (see table S1.A): Women were scored lower than men in research productivity and research impact but higher than men in contributions to diversity. In the other categories, including the overall impression category, scores for men and women were not significantly different.

To determine whether gender bias was incorporated into rubric scores, we analyzed the research productivity category because it can be most directly compared with external metrics. We chose two metrics calculated for the candidate's application year. First, we tallied from candidate curricula vitae the number of articles published (and confirmed this tally in the Web of Science database). Second, we pulled from the Web of Science the H-index, a dominant measure of researcher output that incorporates productivity and impact in a single number that can be compared across faculty of all seniority levels (*9*). We call research productivity, a rubric category that can be measured independently, a "calibration category" (see table S2, footnote).

To test for gender bias, we constructed ordinary least squares (OLS) regression models to predict rubric scores of research productivity, controlling for the independently measured categories of seniority and number of articles or H-index (see table S2). We found that women candidates, on average, received statistically significantly lower productivity rubric scores than those of men, even after controlling for seniority (measured as number of years since PhD) and number of articles published [unstandardized β coefficient (B) = −0.36, $P \leq 0.01$] (see table S2, model 1). Similarly, women receive significantly lower scores on average than men while controlling for seniority and H-index (B= −0.29, P< 09.05) (table S2, model 3). Thus, rubric scoring alone did not appear to fully mitigate gender bias. These findings mirror the gender bias detected in other academic peer-review processes (*10*, *11*). Because the H-index itself has been found to incorporate bias against women (see table S8), our findings should be interpreted as additional bias.

Social psychology literature on double standards finds that among more junior candidates, women are often held to higher standards of competence than men in ways that sometimes change for candidates with more experience (*12*). We thus tested whether the effect of gender on rubric scores is contingent on the value of the external metric (see table S2, models 2 and 4).

Women's productivity rubric scores are consistently below those of men who have the same number of articles and seniority (see the first figure, left). Women face an average 0.36-point penalty, which remains the same across the range of number of articles published. (The rubric scores are mostly clustered at the middle to high portion of the scale, between 2 and 4; within this range, there is an approximately 18% penalty for being a woman.)

Yet when controlling for H-index and seniority, the gender penalty is harshest among candidates with the lowest H-indices (see the first figure, right), who are disproportionately junior (see table S4, footnote). At the lowest tail of the H-index distribution, men receive research productivity rubric scores that are on average 0.7 points higher than the scores of women with the same seniority. At this end of the H-index distribution, the rubric scores are mostly clustered between 1 and 3; for these candidates, there is an approximately 35% penalty for being a woman.

The gender difference in rubric scoring gradually decreases by about 0.01 point for each 1 point of H-index gained. Yet women do not catch up to men in how productive they are rated in the rubrics until reaching an H-index of 17.5, a productivity index well above the 12.8 average and achieved by only a handful of candidates. Because rubric scores for men and women in the overall impression category were not statistically significantly different (see table S1), evaluators may have combined category scores so that women's higher average scores on contributions to diversity offset their lower average scores on productivity and impact.

Indication of gender bias in rubric scores on research productivity

Graphs show predicted values by gender from ordinary least squares regression models regressing rubric scores of research productivity on independent productivity metrics, controlling for seniority. n = 62 semifinalist candidates (see table S2, models 1 and 4).



GRAPHIC: K. FRANKLIN/*SCIENCE*

Content analysis of qualitative comments

Next, we conducted what to our knowledge is the first content analysis of qualitative rubric comments in a real faculty search context. Candidates received written commentary alongside their rubric scores. An average of three and a maximum of nine evaluators wrote comments on each candidate. The number of comments received did not differ by candidate gender.

In our content analysis, we prepared a dataset of comments, in which candidate gender was concealed, by removing gender indicators such as pronouns. We then combined an inductive exploration for emergent themes with deductive searches for specific patterns found in previous literature on letters of recommendation. We conducted hand coding, which some research suggests is superior to computer-assisted coding for studies such as ours with new analyses (*13*).

Many comments contained evaluative notes on the quality, number, authorship order, or impact of the publications. Inductively, we coded these as either negative (for example, "some gaps in the pubs." and "only one paper from a postdoc of three-plus years") or positive (for example, "strong publication record, letters attest to research

output and collaboration"). We also searched deductively for the presence of two themes mentioned in previous literature on other types of evaluation. One code is "standout" language (*13*, *14*). Examples in the rubric comments include "outstanding productivity and quality" and "probably the best [search specialty] candidate out there this cycle." The last code is "doubt raisers" (*15*), when a seemingly positive or neutral comment is accompanied by language that minimizes the accomplishment or raises concerns (for example, "several publications, but... some impact factors are very modest"). We turned these four codes into four dichotomous measures of the presence or absence of comments in each category. (See table S3 for details on coding methods, intercoder reliability, dichotomous measure rationale, and robustness check.)

Using rubrics for EDI hiring

Recommended process for rubric use to improve EDI in faculty hiring. Steps 1, 3, and 6 are conducted by faculty evaluators while steps 2, 4, and 5 are conducted by faculty on the search committee with administrative support.

Evaluators / **Process managers**

1 — Evaluators: Work as a group to determine rubric categories and weights, in accordance with disciplinary schemes of merit, to mitigate evaluation bias.

2 — Process managers: Ensure that rubric includes a calibration category, i.e., one that is independently quantifiable.

3 — Evaluators: Individually complete rubrics for all candidates to mitigate individual bias.

4 — Process managers: Compile and analyze rubric results, comparing calibration category scores to independent metrics to detect any remaining individual bias.

5 — Process managers: Start the meeting of evalutors by presenting rubric results to mitigate first speaker bias, noting high inter-evaluator variance and any large inconsistencies of scores with independent metrics.

6 — Evaluators: Use findings to guide holistic discussion of candidates, futher reducing individual bias and mitigating interactional bias.

GRAPHIC: C. SMITH/*SCIENCE*

Candidate gender was subsequently unmasked, and the percentages of women and men who had received at least one positive comment, negative comment, standout term, and doubt raiser were calculated (see table S3). We found that 86% of men but only 63% of women candidates received at least one positive comment. Men were half as likely to receive a negative comment (25%) compared with women (50%). Men were 3.5 times more likely to receive standout language (32%) compared with women (9%). A χ^2 test indicates statistically significant gender differences for these three variables. Thirteen percent of women and 25% of men received a doubt-raising comment. This pattern was not expected, yet this gender difference is not statistically significant (see table S3). Overall, the gendered patterns in rubric quantitative scores align with the qualitative comments and with previous research on other evaluative language (*13*–*15*).

Survey of faculty

At the conclusion of Phase Two, we conducted an anonymous survey of department faculty, to which 56% of the professors responded. Our check for selection bias in responses revealed no statistically significant difference in the two indicators we have for respondents and nonrespondents: gender and tenure status (see table S6). Most reported that rubric usage helped them evaluate candidates in a more organized fashion (78%) and may have helped them be more objective (78%), potentially reducing individual bias (see fig. S2 and table S5).

At the beginning of the meeting to discuss the semifinalist list and choose finalists to invite for interviews, a faculty member presented rubric numeric scores and comments with the identity of the evaluators anonymized. It was explained to faculty that this step aimed to enable all viewpoints to be heard while avoiding first-speaker bias, in which initial speakers set the tone for discussion. However, faculty were not discouraged from making additional comments or claiming or echoing their support for particular comments after the presentation. Most survey respondents said that this practice prompted meeting attendees to focus on more objective criteria (78%), improved the climate of the meeting (80%), and reduced the first speaker effect (67%) (see fig. S2).

Additionally, some commented in open-ended survey responses that meeting time was used efficiently to quickly identify candidates with strong consensus and then spend more time on those with high variance in ratings [whom we determined were disproportionately women (see table S1.B)]. Taken together, these results suggest that beginning the faculty meeting with rubric results reduced interactional bias emergent in the faculty meeting. By opening with a neutral reading of the full set of both positive and negative rubric comments, the impact was blunted of any first speakers, often senior men, attempting to vociferously promote or shoot down a candidate. The faculty meeting format may have mitigated gender bias in the research productivity scores and the selection of finalists; 47% of women semifinalists and 37% of men semifinalists advanced to the finalist stage.

POLICY TEMPLATE

In light of our findings that gender bias remains endemic even in this seemingly objective evaluation process, it is vital that rubric usage be accompanied by strategic application in departmental meetings to counteract individual bias and check interactional bias during the discussion of candidates. Our results suggest that using rubrics according to this process framework can improve diversity in hiring. Thus, we recommend a strategic and sociologically astute use of rubrics as a department self-study tool within the context of a holistic evaluation of the short-listed candidates (see the second figure).

We have studied this process with regard to gender. Given the otherwise limited diversity among candidates in our study, we were unable to address whether rubrics could also be a tool to promote and check on the fairness of evaluations with regard to race/ethnicity or other minoritized identities. This suggests priorities for future research. ■

REFERENCES AND NOTES

1. M. Blair-Loy, E. Cech, *Misconceiving Merit: Paradoxes of Excellence and Devotion in Academic Science and Engineering* (Univ. Chicago Press, 2022).
2. J. Roy, C. Wilson, A. Erdiaw-Kwasie, C. Stuppard, "Engineering and engineering technology by the numbers 2019" (American Society for Engineering Education, 2020).
3. J. Falkenheim, "Doctoral Recipients from U.S. Universities: 2019" (National Science Foundation, 2019).
4. K. O'Meara, D. Culpepper, L. L. Templeton, *Rev. Educ. Res*. **90**, 311 (2020).
5. University of Michigan, in *Advance Program,* University of Michigan, Ed. (University of Michigan Office of the Provost, 2018).
6. E. Uhlmann, G. L. Cohen, *Psychol. Sci*. **16**, 474 (2005).
7. R. H. Thaler, C. R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale Univ. Press, 2008).
8. National Research Council, *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty* (National Academies Press, 2010).
9. V. Koltun, D. Hafner, *PLOS ONE* **16**, e0253397 (2021).
10. C. Wenneras, A. Wold, *Nature* **387**, 341 (1997).
11. E. R. Andersson, C. E. Hagberg, S. Hägg, *Front. Res. Metr. Anal*. **6**, 594424 (2021).
12. M. Foschi, *Annu. Rev. Sociol*. **26**, 21 (2000).
13. S. J. Correll, K. R. Weisshaar, A. T. Wynn, J. D. Wehner, *Am. Sociol. Rev*. **85**, 1022 (2020).
14. T. Schmader, J. Whitehead, V. H. Wysocki, *Sex Roles* **57**, 509 (2007).
15. J. M. Madera, M. R. Hebl, H. Dial, R. Martin, V. Valian, *J. Bus. Psychol*. **34**, 287 (2019).

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, grant 1661306 (M.B.-L.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. S.I.F. is a cofounder and scientific adviser for MelioLabs and holds equity in the company.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm2329

10.1126/science.abm2329



Tech jobs performed on large corporate campuses like Apple's should be decentralized, argues Khanna.

BOOKS *et al.*

SCIENCE AND SOCIETY

The digital citizen

Two new books advocate for technological agency

By **Michael Neblo**[1,2] and **Avery White**[2]

Many people initially regarded the radical ideas espoused by Charles Darwin about evolution and Henry Ford about automation technology as threats to what was understood to be, respectively, humankind's proper place in the world and dignified work. By adjusting our ideas and practices, however, we have (mostly) been able to reconcile ourselves to the upheavals wrought by the evolutionary worldview and the economics of industrial automation. Two new books, Ro Khanna's *Dignity in a Digital Age* and Jamie Susskind's *The Digital Republic*, aim to help today's society adapt to the social, economic, and political upheavals underway as a result of rapid changes in digital technology.

Khanna—a US congressman whose district includes Silicon Valley—synthesizes a remarkable range of philosophy, economics, tech knowledge, and savvy politics in his plan to reduce economic inequality (part I) and partisan polarization (part II), while preserving people's ability to establish and maintain roots in the communities of their choice. He argues that people in regions without access to high-paying tech jobs often face a real (but ultimately unnecessary) choice between abandoning their communities and foregoing the work and wealth afforded by the tech industry. The indignity of that choice, and the resulting economic inequality, drive much of the recent increase in political polarization, he maintains. This process is further catalyzed by the often insidious effects that social media itself has had on our political discourse. He proposes instead a policy to decentralize the tech industry geographically and an "Internet Bill of Rights" incentivizing industry to promote healthy public discourse. Khanna's vision sometimes gets lost in the litany of proposals he puts forward, but the insight it offers is potentially profound.

Dignity plays a different role for Susskind, who titles his opening chapter "The Indignant Spirit." He grounds his thesis in the "republican" tradition of political philosophy, which emphasizes avoiding subjection to the arbitrary power of others. Citizens become indignant when they are subject to governmental or private power, he argues, and should feel the same when big tech similarly encroaches on their freedom.



Dignity in a Digital Age
Ro Khanna
Simon & Schuster, 2022. 368 pp.



The Digital Republic
Jamie Susskind
Pegasus, 2022. 304 pp.

Although it comes with a "parsimony" caveat, Susskind's "digital republicanism" inverts traditional concerns about state power. The primary threat to be contained, in his view, emanates from private industry, and democratically guided state action is necessary to curtail it.

The private power held by the technology sector is all the more insidious because it often operates covertly, argues Susskind. Coding practices can produce relations of domination, for example, even when coders act in good faith. Even potential coercion, he maintains, is a serious problem—it is not enough to rely on the goodwill of individuals in the tech sector or on market discipline. We must build institutions that afford democratic citizens a robust say in the forces that affect them.

Susskind's thesis is thus more general than Khanna's, and his policy proposals are applicable beyond the United States. His book includes an entire chapter defending an internationalist component of his approach. But reading these two books together further enhances the value of each. While Khanna links our private lives as workers and our public lives as citizens, he offers more detailed policy proposals related to the former. Susskind's proposals, meanwhile, tend to emphasize changes that might be put in place on the public side. The two books provide a full picture of what it would mean to govern technology well.

It is also worthwhile to consider more radical notions of digital spaces. In his recent book, *Reality+*, for example, philosopher David Chalmers argues that virtual entities are no less real for being digital. If we accept this as true, we should regard such spaces as new venues in which to live a good life, rather than as distractions from the "real" world.

Khanna rightly values an individual's ability to establish and maintain roots in a particular place. His strictly physical interpretation of "place," however, undervalues the ways virtual spaces enable experiments in living that keep liberal democracies robust and vibrant. Similarly, Susskind rightly emphasizes the need to avoid technological domination, but, at times, his position de-emphasizes positive human agency—the power to experiment with different ways of living. Efforts to govern the tech industry should also promote technology's power to enable previously undiscovered ways to live lives of dignity. ■

10.1126/science.abq6958

The reviewers are at the [1]Department of Political Science and [2]Institute for Democratic Engagement and Accountability, The Ohio State University, Columbus, OH 43210, USA. Email: neblo.1@osu.edu

PHOTO: NOAH BERGER/REUTERS/ALAMY STOCK PHOTO

BIOCHEMISTRY

Bringing inanimate matter to life

A biochemist probes the origins and vital nature of the Krebs cycle

By **Joseph Moran**

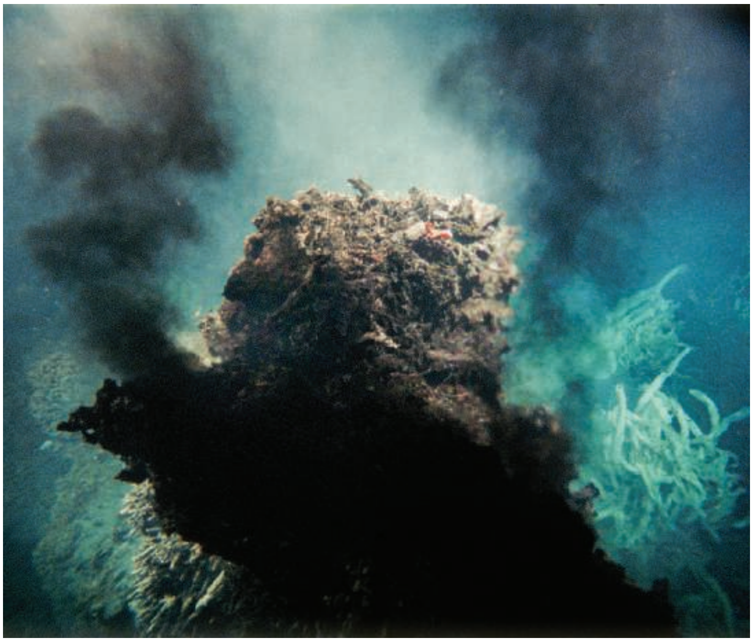
The incredible scientific advances centered around DNA, RNA, and proteins have left many people with the impression that the essence of life is found in genetic information. Yet a cell that died seconds ago contains the same genes that it did moments earlier, suggesting that there are other processes at play that bring the inanimate to life. In his latest book, *Transformer*, biochemist Nick Lane reminds readers in accessible prose that it is life's dynamic chemistry, its metabolism, that draws inanimate matter into the living state and back again. As Lane puts it, metabolism is not just "what keeps us alive—it is what being alive *is*."

The main hero of the book is the Krebs cycle, a sequence of chemical reactions that most of us were taught releases energy from molecules obtained from the breakdown of sugars or fats, expelling CO_2 as a waste product. However, Lane points out that this perspective obscures this vital process's place at the heart of all biology and life itself.

Few laypeople know that the Krebs cycle can run in reverse, fixing CO_2 and hydrogen gas to produce all the key building blocks of biochemistry—a process as natural and thermodynamically favored as water flowing downhill. Some versions of this reverse Krebs cycle contain a feedback loop that amplifies the quantity of its own chemical constituents. Rather than simply being an engine for the synthesis of the molecules of life, it may thus be more accurate to say that life is what one can make from the Krebs cycle.

Lane begins by taking readers on a thrilling tour of the remarkable stories behind the discoveries of some of life's key metabolic pathways and mechanisms. He lays bare the human side of science, telling of pioneering biochemists such as Frederick Gowland Hopkins, Peter Mitchell, Jennifer Moyle, Melvin Calvin, Andrew Benson, and others caught up in intellectual battles with larger-than-life personalities and nearly thwarted by the whims of political currents and (un)fortunate timing. Lane does not feed readers the textbook story, instead wading through the messy process of discovery and the debates that predated the modern consensus over how the central metabolic pathways work.

The reader next learns how chemical forerunners to these biological pathways likely gave rise to life. Experiments from my own lab and conducted by others are now starting to support such ideas (*1*). Lane and others, such as Jack Corliss, Günter Wächtershäuser, Mike Russell, and Bill Martin, speculate that deep-sea hydrothermal vents were life's birthplace. Some of these researchers argue that a primitive form of the Krebs cycle led to the first living systems and that the cycle may have initially run in the reverse direction. The proton gradients between the inorganic membranes found within these vents may have been key to promoting the chemistry of life, argues Lane.



Hydrothermal vents like this one may have facilitated early biochemical processes.



Transformer: The Deep Chemistry of Life and Death
Nick Lane
Norton, 2022. 400 pp.

Moving on to early life, Lane reasons that many of the key evolutionary transitions in deep time can be understood by considering how geological changes or biological innovations affected the Krebs cycle. Subsequent chapters paint a fresh perspective on the centrality of the Krebs cycle to disease, aging, and even what it means to be an individual. Some of these ideas are undoubtedly speculative but are undeniably stimulating—the mark of good popular science writing.

Biased by my own interest in the chemical origins of metabolism, I found myself riveted, especially by the first half of the book. Lane does a masterful job of explaining the needed chemistry in plain language, but I suspect that readers without such a background will find themselves flipping back in places to refresh their memories on chemical names. In a few spots, I could not tell which ideas came from other researchers and which were Lane's—a literary tactic, I presume, to keep things straightforward. Those who want to know to whom each theory and discovery is attributed can peruse the more than 60 pages of annotated references that cap the book.

But these are minor quibbles. With *Transformer*, Lane has managed to write a book about the Krebs cycle that jumps fresh off the page, overturning any residual antipathy for biochemistry one might still be holding on to from his or her student days. The book brings to life the chemistry that brings us to life. ■

The reviewer is at the Institute of Supramolecular Science and Engineering (ISIS), University of Strasbourg, 67000 Strasbourg, France. Email: moran@unistra.fr

REFERENCES AND NOTES

1. K. B. Muchowska, S. J. Varma, J. Moran, *Chem. Rev.* **120**, 7708 (2020).

10.1126/science.abp8953

PHOTO: RALPH WHITE/GETTY IMAGES

INSIGHTS

LETTERS



Pandemic-inspired policies

The COVID-19 pandemic disrupted long-standing routines, creating myriad challenges and opportunities. Now that many restrictions have been lifted, a new normal is taking shape. We asked young scientists to describe positive changes their institutions have made, or should make, based on lessons learned during the pandemic. Read some of the best responses here. Follow NextGen Voices on Twitter with hashtag #NextGenSci. **—Jennifer Sills**

Flexible schedules

My organization has embraced a hybrid work model: Employees can choose to come to the office for only 2 or 3 days a week (or fewer in the case of parents of newborns). As the institution understands the importance of both flexibility and creating a strong network within the team, team-building games are conducted remotely. This strategy helps people to relax in the midst of hectic work, thereby building a healthy and friendly work atmosphere.

Nandhini Kesavan
Experian Services India Pvt. Ltd., Mumbai 400070, India.
Email: nandhinik7@gmail.com

Before the pandemic, all lab members were expected to be present in the lab, even outside working hours and on weekends. People who were present for long hours were considered smart and hard working. The pandemic showed that it is possible to collaborate and work effectively in a remote setup. Now, researchers have the much-welcomed flexibility to work from home when writing an article or thesis.

Senthilkumar Seenuvasaragavan
Mumbai 400076, India.
Email: shrisenjohn@gmail.com

My university has realized that if you are sick, you should stay home, and that this principle applies to everyone—students, staff, and faculty. This is a much kinder and safer policy than the previous expectation that sick people should drag themselves to campus despite their own needs and the risk of contagion.

Katie Burnette
Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, Riverside, CA 92521, USA.
Email: katie.burnette@ucr.edu

Remote learning options

My department has continued to use a hybrid format for departmental seminars and colloquia. Hybrid seminars are accessible; they allow for people to partake in the community-building, in-person format that was common before the pandemic but also offer an opportunity for those who cannot physically be there to stay involved and informed. Even when the speaker is presenting remotely, my department often offers a space for people to view seminars in person. Hybrid seminars can also diversify the speakers. By inviting remote speakers without needing to consider travel logistics, we can greatly

ILLUSTRATION: ROBERT NEUBECKER

broaden the demographics and expertise of our presenters.

Ashley Barbara Heim
Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA. Email: abh229@cornell.edu

Before the pandemic, few professors recorded their lectures or offered synchronous livestreams. The normalization of video streaming and recording applications has offered students who have to miss class the opportunity to stay on track. As a result, students struggling to attend for any reason are less likely to fall behind, and students who are sick know that they can stay home to recover, improving their health and protecting their classmates.

Rishi Jai Patel
Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: ripatel@sas.upenn.edu

Financial support

During the pandemic, our department learned to be more responsive to students in financial crisis. When a remote student's laptop was stolen (and shipping a library loaner device wasn't feasible), we confirmed that they were eligible for a grant to replace the equipment. Immediately, the department provided cash for the purchase, allowing the student to buy a new laptop and return to class. Weeks later, when the grant was fully approved, the department received a reimbursement for the expense. This quick response demonstrated that prompt financial support is possible.

Juliet Johnston
Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA 94720, USA. Twitter: @queermsfrizzle

My PhD institution, Cornell University, offered continued financial support to international graduate students who were stranded in their home countries due to government-imposed travel restrictions. The support included options such as graduate teaching assistantship (for online courses), graduate research assistantship (for studies that could be done remotely), and support for students to attend online conferences. This innovative change made it possible for international students to maintain progress toward their research degrees during the COVID-19 disruptions. Some students also had the opportunity to conduct research with collaborators of their advisers at institutions in the students' home countries. Providing international students with the option to temporarily continue their work remotely could alleviate stress for those who need to travel home in the future.

Xiangkun (Elvis) Cao
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Twitter: @Elvis_Cao

Mental health services

The pandemic lockdowns began shortly after I arrived in China to begin my postdoc. I was isolated in a new place, far from family, and my mental health suffered. My institution could have helped by providing psychiatrist consultations for foreign postdocs. With more support, I could have adjusted more quickly and been more productive. Such services should always be available to those experiencing stress, especially foreign students and employees.

Hafiz Arbab Sakandar
Food Engineering Department, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia, China. Email: hafizarbabsakandar@gmail.com

After the pandemic lockdowns, my institution began to pay more attention to the mental health of students, especially those who faced stalled projects or difficulties finding a job. I was one of the "Psychological Committee" volunteers, recruited by the Students Counseling and Mental Health Center. We organized activities to help students relieve stress and make friends. The committee also answered mental health questions on social media. During this time, I received training from the university on how to recognize and respond to psychological crises, which helped me support my classmates. The university has also made professional psychological consultants available to students. Both the services offered and the training I received should be made widely available going forward.

Yan Zhuang
Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China. Email: zhuangyanisme@pku.edu.cn

Facilitation of exercise

My university sports center adapted to pandemic concerns by starting an electronic sign-up and sign-in system to use the facilities. At first a way to limit capacity and conduct contact tracing, the system is still in place today. Some may find the requirements burdensome, given that walk-in attendance was possible before the pandemic and there are penalties for not showing up. However, for me, this change promotes accountability, adds structure to the day, and has led to better time management.

Kathryn K. Oi
Department of Chemistry, University of Zurich, 8057 Zürich, Switzerland. Email: kathryn.oi@chem.uzh.ch

My organization has established a mobile sports platform for employees and students to plan personal fitness and exercise. With this software, people can make personalized indoor exercise training plans and view professional guidance at home. The benefits of helping employees meet their fitness goals will long outlast the pandemic.

Yuan Zhi
School of Economics, Guizhou University, Guiyang, Guizhou 550025, China. Email: yzhi@gzu.edu.cn

Online tools

Before the pandemic, most faculty at my university didn't know how to use online teaching software, and the university didn't have an online teaching resource database. During the pandemic, the office of academic affairs conducted systematic training for teachers and students, made a series of teaching videos, and established corresponding websites for students to browse and study. This online teaching platform provides crucial resources that both faculty and students continue to use.

Yujie Diao
Department of Blood Transfusion, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui 230022, China. Email: dyjahhf@163.com

Because of the pandemic, digital innovations in higher education that would have otherwise taken years to implement were introduced in a matter of days. The changes were examples of "creative destruction," a concept originated by economist Joseph Schumpeter. My institution started using educational technology driven by artificial intelligence that predicts the navigational behavior of learners in an online course and identifies pedagogically valuable behaviors in real time. The technology serves as an online tutor that can assess a student's strengths and weaknesses and deliver personalized individual instructions, extending the capacity of a classroom teacher.

Wadim Strielkowski
Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA 94720, USA. Email: strielkowski@cantab.net

10.1126/science.add2046

RESEARCH



IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

NEURODEVELOPMENT

A model for dendrite branching

Highly complex branched dendritic arbors allow neurons to integrate many inputs, but it has been difficult to analyze how such variable structures form. Shree *et al*. combined live imaging of the dendrites of sensory neurons in *Drosophila* larvae with computational modeling to derive the rules that drive their development. A model in which the tips of dendrites randomly switch between growing, shrinking, and pausing can reproduce the observed patterns of growth. The model also reveals the importance of branching for increasing the size of dendritic trees and identifies properties that may give different types of neurons their morphologies. —JET *Sci. Adv.* 10.1126/sciadv.abn0080 (2022).

Fluorescence microscopy image of a fruit fly larva revealing the branching patterns of sensory neuron dendrites

IMAGE: SHREE *ET AL.*

EVOLUTION

From one mutation, many effects

New mutations introduce genetic variation into a population, and if the genetic changes persist, they can provide raw material for evolutionary change. Pleiotropy, or the number of traits altered by a mutation, is often assumed to affect the probability that a mutation survives the test of time. Vande Zande *et al*. now provide a framework with which to test such assumptions by empirically determining and comparing the pleiotropic effects of cis- and trans-regulatory mutations affecting expression of the *TDH3* gene in yeast. They found that trans-regulatory mutations do indeed have more pleiotropic effects and are more deleterious. —LMZ and DJ

Science, abj7185, this issue p. 105

NUTRIENT METABOLISM

Leucine sensing in the zone

Leucine is an essential amino acid that plays an important role in a variety of metabolic processes, including both physiological ones such as insulin secretion and pathological ones involved in the metabolic syndrome. mTORC1 is a key protein that is also involved in many aspects of metabolism and senses leucine through proteins called sestrins. Some aspects of this signaling pathway were previously observed in cultured cells, but now Cangelosi *et al*. have identified what happens in vivo in a mammal. The authors clarified the signaling steps involved in mTORC1 sensing of leucine through sestrins in mice and determined that the regulation of this pathway varies between different zones within the liver based on proximity to blood vessels. —YN

Science, abi9547, this issue p. 47

NANOMATERIALS

Trapping and recovering organics

Hydrogels consist of cross-linked organic polymers that can swell to hold up to 90% water, making them useful as absorbents and for tissue engineering. Zhang *et al*. synthesized inorganic nanowires from polyoxometalates, calcium ions, and oleylamine and found that these nanowires readily formed three-dimensional networks. The networks swell when exposed to a range of volatile organic compounds added at fractions even below 1% to form organogels. The gels are stable to physical squeezing without a substantial loss of liquid. However, the liquids can be recovered using distillation and centrifugation, and the nanowires can be reused, making possible the trapping and recovery of organic solvents. —MSL

Science, abm7574, this issue p. 100

HUMAN GENOMICS

Human migrations into Micronesia

The movements of ancient humans can be difficult to ascertain from their current population genetic structure. Studying the peopling of the Micronesian islands, Liu *et al*. examined 164 ancient human remains from five different archaeological sites in remote Oceania from different prehistoric time frames, along with 112 present-day individuals from the same area. They combined these new data with the results of previous studies and also compared their results with linguistic studies. Their analysis revealed successive movements from island Southeast Asia that differ from those in the southwest Pacific. Furthermore, co-analysis of Micronesian and southwest Pacific ancient DNA indicates that the first people who colonized the Pacific islands had a population

structure in which men moved to find their mates, whereas women rarely moved to join men. —LMZ

Science, abm6536, this issue p. 72

PHYSIOLOGY

Balancing liver repair and regeneration

The related cytokines transforming growth factor-β (TGF-β) and bone morphogenic protein (BMP) can exert protective or detrimental effects during liver injury. Stavropoulos *et al*. investigated the contributions of the TGF-β and BMP pathways in the response to acetaminophen-induced liver injury in reporter-expressing mice. Both pathways were induced near areas of necrosis in hepatocytes that also showed an increase in autophagy, and both pathways were required to stimulate autophagy. Although inhibiting signaling through both pathways initially worsened the damage, it ultimately accelerated tissue regeneration and recovery. —AMV

Sci. Signal. **15**, eabn4395 (2022).

PALEOCLIMATE

Warmer than realized

Past climates may hold important clues about how the planet might respond to ongoing climate warming. Meckler *et al*. use clumped isotope thermometry on benthic foraminifera to reinterpret the record of the deep ocean temperature over the past 65 million years. They found that deep ocean temperatures were generally higher and more variable than earlier work suggests. Their results have implications for our understanding of deep sea temperature responses to atmospheric carbon dioxide concentrations, climate sensitivity, and ocean structure during times of minimal continental ice. —HJS

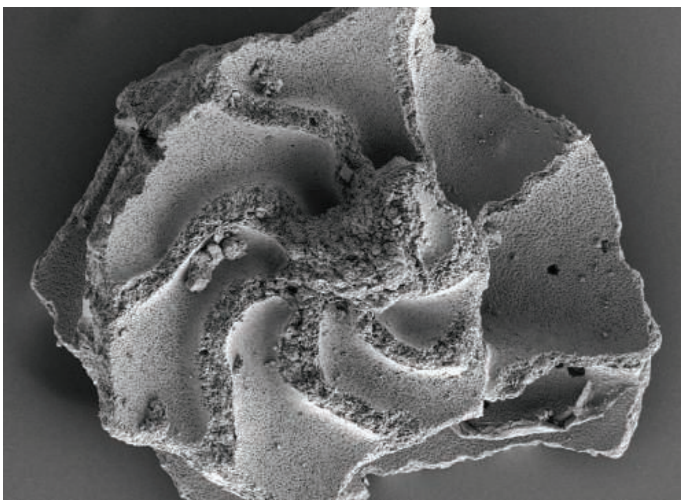
Science, abk0604, this issue p. 86

NEUROSCIENCE

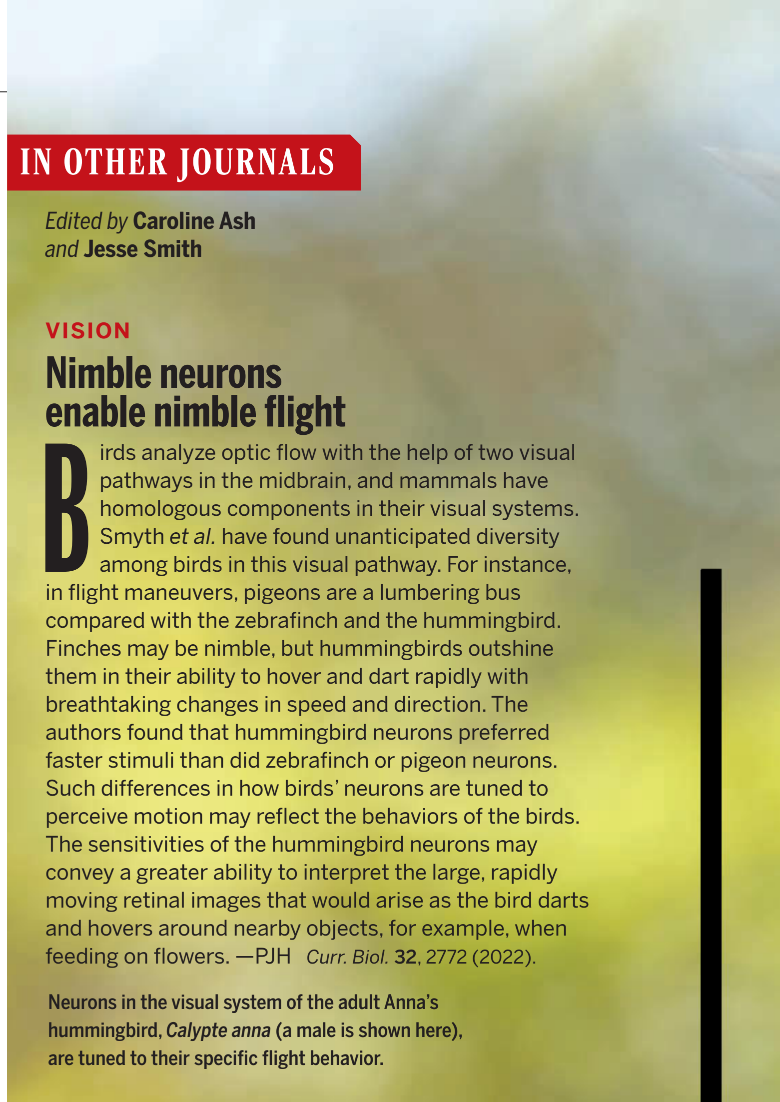
Microglia can cause chronic pain

Peripheral nerve injury leads to long-lasting pain hypersensitivity. Treatment of chronic pain continues to be unsatisfactory because we still don't fully understand the underlying processes. Tansley *et al*. discovered a new mechanism by which peripheral nerve injury causes pain. A subpopulation of projection neurons in the spinal cord dorsal horn that are critical for transmission of pain signals to the brain, but not other cell types in superficial dorsal horn, are selectively surrounded by specialized extracellular matrix structures called perineuronal nets. Nerve injury–activated microglia degrade perineuronal nets around projection neurons, which increases projection neuron activity, subsequently causing pain. —PRS

Science, abl6773, this issue p. 80



Fossil foraminifera, seen here with scanning electron microscopy, can be analyzed to determine the ocean temperature millions of years ago.



IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

VISION

Nimble neurons enable nimble flight

Birds analyze optic flow with the help of two visual pathways in the midbrain, and mammals have homologous components in their visual systems. Smyth *et al.* have found unanticipated diversity among birds in this visual pathway. For instance, in flight maneuvers, pigeons are a lumbering bus compared with the zebrafinch and the hummingbird. Finches may be nimble, but hummingbirds outshine them in their ability to hover and dart rapidly with breathtaking changes in speed and direction. The authors found that hummingbird neurons preferred faster stimuli than did zebrafinch or pigeon neurons. Such differences in how birds' neurons are tuned to perceive motion may reflect the behaviors of the birds. The sensitivities of the hummingbird neurons may convey a greater ability to interpret the large, rapidly moving retinal images that would arise as the bird darts and hovers around nearby objects, for example, when feeding on flowers. —PJH *Curr. Biol.* **32**, 2772 (2022).

Neurons in the visual system of the adult Anna's hummingbird, *Calypte anna* (a male is shown here), are tuned to their specific flight behavior.

SIGNAL TRANSDUCTION

Signaling the bacterial vote

Quorum sensing, the mechanism by which bacteria monitor the size of their population by responding to signaling molecules (autoinducers) produced by individual cells, may resemble more of a community vote than a simple count of the number of bacteria. Bettenworth *et al.* found that the bacterium *Sinorhizobium meliloti* produces SinI, the enzyme that makes its autoinducer (a homoserine lactone), in a stochastic, pulsatile manner. The Sin regulatory system acts as a probabilistic switch that activates relatively rarely but is influenced by the cell's response to physiological cues such as growth conditions, which can change the pulse frequency. A quorum-sensing response occurs, depending on the combined contribution of bacterial cells to the common pool of autoinducer molecules. —LBR

Nat. Commun. **13**, 2772 (2022).

ELECTROCHEMISTRY

Toward CO_2-free ammonia synthesis

Lithium-mediated electrochemical ammonia synthesis (LiMEAS) is thought to be a promising alternative to the traditional thermochemical Haber-Bosch process because it can operate at ambient temperature and pressure conditions and without CO_2 emissions. However, little is known about its complex mechanism. Using standardized high-resolution gas chromatography–mass spectrometry and nuclear magnetic resonance imaging, Sažinas *et al.* analyzed the stabilizing effect of oxygen in LiMEAS. They determined the

CREDITS: (LEFT TO RIGHT) MECKLER *ET AL*.; DAN SUZIO/SCIENC

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

GUT PHYSIOLOGY

Regulating the redox swamp

The diversity and abundance of organisms that live in the human gut are well known. We also have some idea of associations between microbial taxa and host disease, but we have failed to consistently translate this understanding into therapeutic options. Redox reactions, which are governed by the availability of respiratory electron acceptors, are the drivers of bacterial community composition. Lee *et al*. reviewed the literature on bacterial colonization of regions of the gut and concluded that colonization is primarily regulated by host physiology controlling electron acceptor availability rather than by microbial activity. Although diet shapes microbiota composition, its effects are largely mediated by its effect on host homeostasis. Intervening with host gut physiology could thus offer a more tractable route to translational solutions for gut dysbiosis than trying to manipulate the microbiota. —CA

Science, abp9960, this issue p. 44

NEUROGENOMIC IMAGING

Mouse and human cortex cell comparisons

Studies of specific cellular types and their molecular phenotype in the cortex have been especially difficult in humans. Fang *et al*. applied an imaging technique called MERFISH to 4000 gene transcripts to identify the spatial distributions of cell types within the human and mouse cortex. From this survey, they identified cell-type-specific association patterns that differed between the mouse and human brain. They found that humans have more specific neurons than mice, and that these cells have greater proximity to and interactions with similar neuronal cell types. Studying receptor-ligand pairs, the authors identified likely contacts between neurons and microglia, which may explain why certain genes are associated with neurodegenerative diseases. —LMZ and PJH

Science, abm1741, this issue p. 56

NEUROSCIENCE

Sleep as a response to stress

Stress-induced sleep has been hypothesized to dampen stress responses and to help mental processing of the stressful events. However, the mechanisms underlying these phenomena are not understood. Yu *et al*. discovered a specific circuit in the mouse midbrain dedicated to detecting stress and inducing restorative sleep (see the Perspective by Joëls and de Kloet). A subset of specialized neurons receive stress inputs and induce rapid eye movement (REM) and non-REM sleep through the lateral hypothalamus. Transient stress enhanced the activity of these neurons for several hours. Stress-induced sleep generated by this pathway alleviated stress levels and mitigated stress-induced anxiety, restoring mental and body functions. Sleep after stress can thus profoundly reduce the chronic stress response. —PRS

Science, abn0853, this issue p. 63; see also adc9782, p. 28

NATURAL HAZARDS

Going on the lamb

The Hunga Tonga undersea volcanic eruption was one of the most powerful ever recorded, with audible sound detected more than 10,000 kilometers from the source. Matoza *et al.* present infrasound and seismic recordings, along with other geophysical observations, that help to describe this event. An atmospheric lamb wave, characteristic of energetic atmospheric events, circled the planet four times and was similar to the 1883 Krakatau eruption. Kubota *et al.* detail how this lamb wave contributed to the global tsunami waves arriving much earlier than expected. The eruption also generated long-range infrasounds and ionospheric interations, along with a global tsunami. This set of observations will be helpful for disentangling the event and understanding the propagation of waves through the atmosphere and ocean (see the Perspective by Brodsky and Lay). —BG

Science, abo4364, abo7063, this issue p. 91, p. 95; see also abq5392, p. 30

MUSCULAR DYSTROPHY

Macrophages modulate muscular dystrophy

The role of resident macrophages in skeletal muscle regeneration after injury has previously been difficult to study. Babaeijandaghi *et al.* used parabiosis, lineage tracing, and single-cell RNA-sequencing experiments to identify a population of self-renewing resident macrophages that were found to clear damage-induced apoptotic cells early after muscle injury. In a mouse model of Duchenne muscular dystrophy, depletion of macrophages through long-term colony-stimulating factor 1 receptor (CSF1R) inhibition resulted in a shift in muscle fiber composition toward more glycolytic-oxidative fibers that were protective against contraction-induced injury. These findings highlight the contribution of resident macrophages to skeletal muscle regeneration and suggest a role for CSF1R inhibition in the treatment of muscular dystrophies. —MN

Sci. Transl. Med. **14**, eabg7504 (2022).

CORONAVIRUS

A new normal?

The implementation of nonpharmaceutical interventions to mitigate the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has also affected the transmission and epidemiology of various other endemic infectious diseases that are spread from person to person. Because human contacts remain different from what they were before the COVID-19 pandemic, it is possible that these endemic diseases may also adapt to changes in social contacts. This may result in changes in seasonality, and other infections may become less or more severe as susceptible age groups change. In a Perspective, Nelson and Lopman discuss the impacts of reduced social contacts on the transmission of endemic infectious diseases and whether there is an opportunity to eradicate them altogether. —GKA

Science, abp9316, this issue p. 33

BIOCHEMISTRY

Keeping things dynamic

Many proteins contain domains that do not adopt a stable folded structure but are involved in protein function, sometimes through self-association that can drive the formation of phase-separated structures inside cells. These regions are known as low-complexity (LC) domains because they are composed of only a few different types of amino acids. Zhou *et al*. show that the LC domain of the RNA-binding protein TDP-43 associates transiently through a labile structure, and this behavior is required for phase separation, which is key to its normal role of organizing mRNA (see the Perspective by Petsko and Small). Mutation of a proline in the LC domain caused aggregation by stabilizing the labile structure. In three other proteins, proline mutations linked to neuronal diseases were shown to similarly enhance the stability of self-association of LC domains. —VV

Science, abn5582, this issue p. 46; see also adc9969, p. 31

GENE REGULATION

HoxA synthetic regulatory reconstitution

Regulatory inputs from many elements are integrated across large genomic windows to control mammalian gene expression. Despite advances in gene editing, generating combinatorial variant alleles to precisely measure the contribution of each input remains challenging. Inspired by the bottom-up approaches of synthetic biology and biochemical reconstitution, Pinglay *et al*. have established a "synthetic regulatory reconstitution" framework for the study of gene regulation. They leveraged advances in large DNA construct synthesis and cell differentiation to dissect the relative contributions of the many mechanisms controlling mammalian *Hox* gene expression. This study opens the door to similar analyses of other loci in any genome. —BAP and SMH

Science, abk2820, this issue p. 45

PALEOCLIMATE

Crustal effects on climate

Why was the long-term global cooling trend of the Cenozoic interrupted by a several-million-year interval of warming during the middle of the Miocene? Herbert *et al*. present a reconstruction of global ocean crustal production to show that tectonic degassing of carbon can account for most of the long-term ice sheet and global temperature evolution for the past 20 million years (see the Perspective by von der Heydt). These results provide further support for the idea that sea floor spreading rates can control global changes in climate. —HJS

Science, abl4353, this issue p. 116; see also abq6542, p. 28

CANCER IMMUNOLOGY

Deleterious release

The secretion of senescence-associated secretory phenotype (SASP) factors by senescent hepatic stellate cells promotes the development of hepatocellular carcinoma (HCC), but the release mechanism is not well understood. Yamagishi *et al.* used an obesity-induced HCC mouse model to study SASP factor release from hepatic stellate cells within tumors. Interleukin-33 (IL-33) release was triggered by lipoteichoic acid, which accumulated in the livers of obese mice. Full-length IL-33 was cleaved into its short, active form and released into the tumor microenvironment through gasdermin D N-terminal pores. Secreted IL-33 bound to $ST2^+$ regulatory T cells, resulting in impaired $CD8^+$-mediated antitumor immunity and the development of HCC. These findings shed light on SASP factor release and suggest that targeting IL-33 release from senescent cells may be an attractive therapy for HCC. —HMI

Sci. Immunol. **7**, eabl7209 (2022).

GPCR SIGNALING

Cholesterol's spot in chemokine receptors

The activation of many G protein–coupled receptors (GPCRs) is allosterically modulated by endogenous lipid molecules, including cholesterol; however, the corresponding molecular mechanisms remain elusive. Using cryo–electron microscopy, Lu *et al.* determined apo- and ligand-bound structures of the signaling complex between a chemokine receptor, a representative of the CX3C subfamily, and its cognate G_i protein. A distinct activation mode of the receptor was identified, along with three cholesterol molecules that stabilize the active receptor conformation and modulate signaling. Structural features of both receptor and ligand were found to define selectivity, in contrast to the largely promiscuous interactions between most chemokine receptors and chemokines. —VC

Sci. Adv. 10.1126/sciadv.abn8048 (2022).

BIOMEDICINE

Cooling away the pain

After injury due to an accident or medical procedure, various forms of pain relief may be required. These can include analgesic medications or local injections to dull the pain receptors, but may also be as simple as applying something cold to the location causing the pain, such as ice packs for sore or bruised joints or muscles. Reeder *et al*. developed a soft, miniaturized, implantable cooler to temporarily block nerve conduction using a liquid-to-gas phase transition as the cooling mechanism (see the Perspective by Jiang and Hong). They borrowed the design of electrical nerve cuff and substituted electrical wires with a microfluidic channel carrying a microliter volume of bioinert coolant. A thermal thin-film sensor integrated within the cuff enabled monitoring of the temperature in real time, thus enabling closed-loop control. —MSL

Science, abl8532, this issue p. 109; see also abm8159, p. 28

structure in which men moved to find their mates, whereas women rarely moved to join men. —LMZ

Science, abm6536, this issue p. 72

PHYSIOLOGY

Balancing liver repair and regeneration

The related cytokines transforming growth factor-β (TGF-β) and bone morphogenic protein (BMP) can exert protective or detrimental effects during liver injury. Stavropoulos *et al*. investigated the contributions of the TGF-β and BMP pathways in the response to acetaminophen-induced liver injury in reporter-expressing mice. Both pathways were induced near areas of necrosis in hepatocytes that also showed an increase in autophagy, and both pathways were required to stimulate autophagy. Although inhibiting signaling through both pathways initially worsened the damage, it ultimately accelerated tissue regeneration and recovery. —AMV

Sci. Signal. **15**, eabn4395 (2022).

PALEOCLIMATE

Warmer than realized

Past climates may hold important clues about how the planet might respond to ongoing climate warming. Meckler *et al*. use clumped isotope thermometry on benthic foraminifera to reinterpret the record of the deep ocean temperature over the past 65 million years. They found that deep ocean temperatures were generally higher and more variable than earlier work suggests. Their results have implications for our understanding of deep sea temperature responses to atmospheric carbon dioxide concentrations, climate sensitivity, and ocean structure during times of minimal continental ice. —HJS

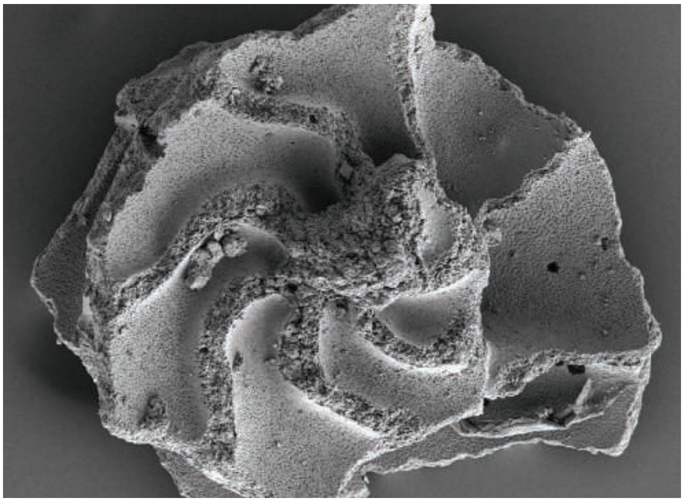
Science, abk0604, this issue p. 86

NEUROSCIENCE

Microglia can cause chronic pain

Peripheral nerve injury leads to long-lasting pain hypersensitivity. Treatment of chronic pain continues to be unsatisfactory because we still don't fully understand the underlying processes. Tansley *et al*. discovered a new mechanism by which peripheral nerve injury causes pain. A subpopulation of projection neurons in the spinal cord dorsal horn that are critical for transmission of pain signals to the brain, but not other cell types in superficial dorsal horn, are selectively surrounded by specialized extracellular matrix structures called perineuronal nets. Nerve injury–activated microglia degrade perineuronal nets around projection neurons, which increases projection neuron activity, subsequently causing pain. —PRS

Science, abl6773, this issue p. 80



Fossil foraminifera, seen here with scanning electron microscopy, can be analyzed to determine the ocean temperature millions of years ago.



IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

VISION

Nimble neurons enable nimble flight

Birds analyze optic flow with the help of two visual pathways in the midbrain, and mammals have homologous components in their visual systems. Smyth *et al.* have found unanticipated diversity among birds in this visual pathway. For instance, in flight maneuvers, pigeons are a lumbering bus compared with the zebrafinch and the hummingbird. Finches may be nimble, but hummingbirds outshine them in their ability to hover and dart rapidly with breathtaking changes in speed and direction. The authors found that hummingbird neurons preferred faster stimuli than did zebrafinch or pigeon neurons. Such differences in how birds' neurons are tuned to perceive motion may reflect the behaviors of the birds. The sensitivities of the hummingbird neurons may convey a greater ability to interpret the large, rapidly moving retinal images that would arise as the bird darts and hovers around nearby objects, for example, when feeding on flowers. —PJH *Curr. Biol.* **32**, 2772 (2022).

Neurons in the visual system of the adult Anna's hummingbird, *Calypte anna* (a male is shown here), are tuned to their specific flight behavior.

SIGNAL TRANSDUCTION

Signaling the bacterial vote

Quorum sensing, the mechanism by which bacteria monitor the size of their population by responding to signaling molecules (autoinducers) produced by individual cells, may resemble more of a community vote than a simple count of the number of bacteria. Bettenworth *et al.* found that the bacterium *Sinorhizobium meliloti* produces SinI, the enzyme that makes its autoinducer (a homoserine lactone), in a stochastic, pulsatile manner. The Sin regulatory system acts as a probabilistic switch that activates relatively rarely but is influenced by the cell's response to physiological cues such as growth conditions, which can change the pulse frequency. A quorum-sensing response occurs, depending on the combined contribution of bacterial cells to the common pool of autoinducer molecules. —LBR

Nat. Commun. **13**, 2772 (2022).

ELECTROCHEMISTRY

Toward CO_2-free ammonia synthesis

Lithium-mediated electrochemical ammonia synthesis (LiMEAS) is thought to be a promising alternative to the traditional thermochemical Haber-Bosch process because it can operate at ambient temperature and pressure conditions and without CO_2 emissions. However, little is known about its complex mechanism. Using standardized high-resolution gas chromatography–mass spectrometry and nuclear magnetic resonance imaging, Sažinas *et al.* analyzed the stabilizing effect of oxygen in LiMEAS. They determined the

CREDITS: (LEFT TO RIGHT) MECKLER *ET AL.*; DAN SUZIO/SCIENC



optimal amount of oxygen to enhance the chemical stability of the system and reported the underlying mechanism. This work is potentially an important step toward a more efficient and stable LiMEAS process. —YS

J. Phys. Chem. Lett. **13**, 4605 (2022).

SOCIAL INEQUITY

Justice, housing and health

Social inequities that place women and racially marginalized renters at disproportionate risk for evictions are associated with HIV risk. However, there are no longitudinal studies that identify the pathways through which evictions affect sexual partnerships that raise HIV risk. Analyzing participants from the longitudinal Justice, Housing and Health Study that were HIV-negative at baseline, Groves *et al.* found that participants who had experienced a landlord-related forced move later reported higher levels of intimate partner violence (IPV) in their sexual relationships during a 6-month follow-up. Furthermore, higher IPV levels were associated with behaviors that increased HIV risk (e.g., unprotected or transactional sex). Policy makers should take note of the results of this study because they reveal how eviction can also affect health by altering the dynamics of sexual partnerships. —EEU

Soc. Sci. Med. **305**,115030 (2022).

NEUROSCIENCE

Emotional arousal promotes memory

Emotional arousal induces noradrenergic activation of the basolateral amygdala, which promotes long-term information storage in other brain regions. The anterior and posterior insular cortex are involved in various forms of recognition memory. After examining basolateral amygdala projections to the insular cortex, Chen *et al.* found that the basolateral amygdala sends dense monosynaptic projections to the anterior insular cortex and only sparse projections to the posterior insular cortex. Memory-enhancing norepinephrine administration selectively reduced neuronal activity within the anterior insular cortex during the postlearning consolidation period but not in the posterior insular cortex. Inhibition of the anterior insular cortex enhances the consolidation of memory, protecting it by disrupting attentional reorienting toward new stimuli and thus promoting ongoing long-term memory storage. —PRS

Proc. Natl. Acad. Sci. U.S.A. **119**, e2203680119 (2022).

POLITICAL SCIENCE

A widening ideological gap

Using three unique decade-long surveys, Jessee *et al.* examined the ideological positioning of the Supreme Court of the United States (SCOTUS) and the perception of the Court by the US public. The authors show that SCOTUS has become much more conservative on key policy issues relative to the general public since 2020. Furthermore, although the US public has adjusted their perception of the ideological shift within SCOTUS, this adjustment on average is less than the actual shift. The researchers also found that the respondents' perception of the SCOTUS ideology relative to their own was correlated with their support for institutional changes, but with noticeable differences between those who identified as Democrats versus those who identified as Republicans. —YY

Proc. Nat. Acad. Sci. U.S.A. **119**, e2120284119 (2022).

PEROVSKITES

Examining edge states

The layered halide perovskites (LHPs) that are used in solar cells can undergo strain relaxation through surface reorganization and form lower-energy edge states (ESs) with higher charge carrier lifetimes and lower emission energies. Qin *et al.* showed that a family of two-dimensional Ruddlesden-Popper LHPs formed $PbBr_6$ octahedra rotated at the crystal layer edges. These dangling octahedra were the origin of localized, lower-energy ESs that helped to transport electrons out of the interior. The population of electrons in both the interior states and ESs could be controlled with external electric fields. Phonons activated by the dangling octahedra interacted with electrons and helped to facilitate radiative recombination, unlike midgap states in conventional semiconductors, which normally quench luminescence. —PDS

Adv. Mater. 10.1002/adma.202201666 (2022).

RESEARCH ARTICLE SUMMARY

GENE REGULATION

Synthetic regulatory reconstitution reveals principles of mammalian *Hox* cluster regulation

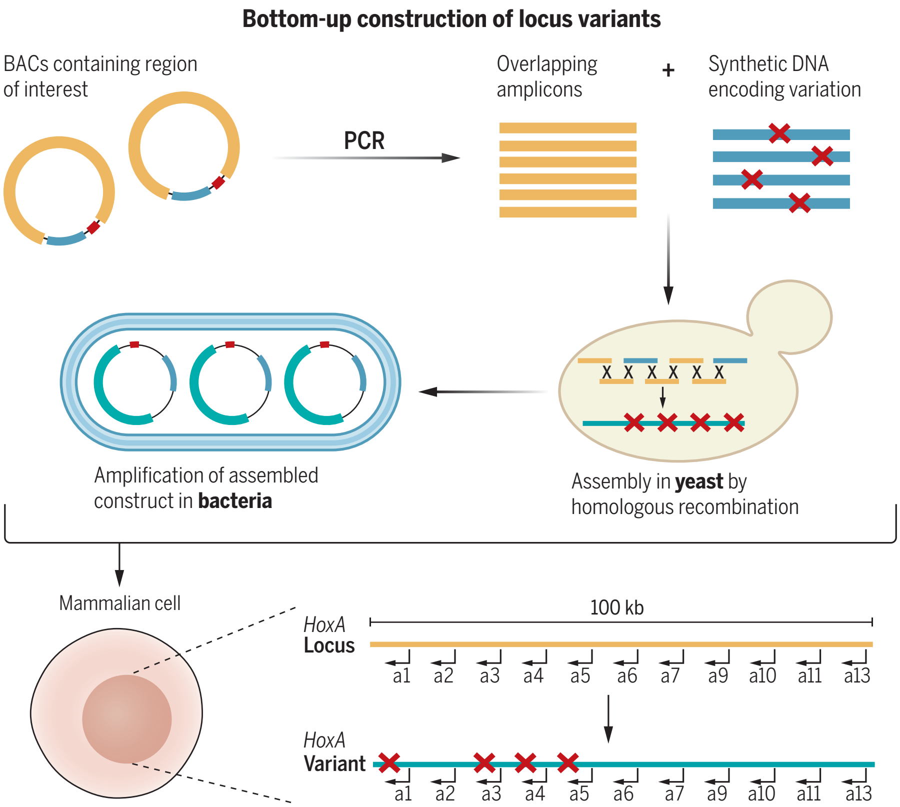
Sudarshan Pinglay†, Milica Bulajić†, Dylan P. Rahe, Emily Huang, Ran Brosh, Nicholas E. Mamrak, Benjamin R. King, Sergei German, John A. Cadley, Lila Rieber, Nicole Easo, Timothée Lionnet, Shaun Mahony, Matthew T. Maurano, Liam J. Holt, Esteban O. Mazzoni*, Jef D. Boeke*

INTRODUCTION: Despite an ever-expanding catalog of noncoding elements that are implicated in the control of mammalian gene expression, how the regulatory input from multiple elements is integrated across a genomic neighborhood has remained largely unclear. This challenge is exemplified at *Hox* clusters (~100 to 200 kb), which contain genes that specify positional identity along the anterior-posterior axis of the developing embryo. In response to developmental morphogens such as retinoic acid (RA), the *HoxA* cluster splits into an active (*Hoxa1–5*) and inactive domain (*Hoxa6–13*) at the level of gene expression and chromatin. Although distal enhancers, intracluster transcription factor binding, and topological organization have emerged as the major regulatory modules in establishing this expression pattern, their relative contributions remain elusive.

RATIONALE: Despite the advent of a vast suite of genome editing tools, it has remained challenging to simultaneously manipulate multiple regulatory elements across large genomic windows to deconvolve their relative contributions. Taking inspiration from the bottom-up approaches of synthetic biology and biochemical reconstitution, we developed "synthetic regulatory reconstitution" as a framework for the study of gene regulation to address this gap. The synthesis of large DNA constructs (>100 kb) permits any combination of complex modifications to be made, at the scale required to probe regulation across a native genomic neighborhood. We fabricated *HoxA* cluster variants that encode various combinations of the previously identified regulatory modules and integrated them into an ectopic location in the mouse genome. This enabled us to directly test the independent ability of these variant ectopic clusters to reconstitute distinct aspects of *HoxA* regulation.

RESULTS: We harnessed the efficient homologous recombination machinery in yeast to construct four rat *HoxA* variants (130 to 170 kb) and delivered them at single copy to the housekeeping *Hprt1* locus of mouse embryonic stem cells. Upon RA-induced differentiation, an ectopic *HoxA* cluster lacking distal enhancers (*SynHoxA*) induced both the appropriate subset of *HoxA* genes and the corresponding chromatin boundary. The presence of distal enhancers (*Enhancers+SynHoxA*) increased transcription levels, especially at early time points. Further, both *SynHoxA* and *Enhancers+SynHoxA* reorganized into active and inactive topological domains upon differentiation, mirroring the endogenous organization. The mutation of just four retinoic acid response elements (RAREs) present in *SynHoxA* (*RAREΔ*) almost completely eliminated any response of the ectopic cluster to RA, at the levels of both gene expression and chromatin reorganization. The addition of distal enhancers to *RAREΔ* could not fully rescue this loss of gene expression phenotype.

CONCLUSION: Our data suggest that at *HoxA*, the primary module of active gene and chromatin boundary specification in response to RA is through the presence of internal transcription factor binding sites. Distal enhancers are dispensable for the specification of active genes but synergize with intracluster activator binding to boost the amount of transcription. Therefore, mammalian *Hox* clusters contain all the regulatory information that is necessary to convert a morphogenetic signal into a stable transcriptional, epigenetic, and topological state. This study showcases the power of synthetic regulatory reconstitution, a generalizable platform for the dissection of gene regulation at other loci in complex genomes. ■



Bottom-up construction of variant mammalian genomic loci at the >100-kb scale. Making multiple edits on the same allele over large genomic windows has remained challenging in mammalian cells. Variant loci with an arbitrary number of changes can be constructed in yeast using synthetic DNA and site-specifically integrated into the genome of mammalian cells to study their behavior. BAC, bacterial artificial chromosome.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: eom204@nyu.edu (E.O.M.); jef.boeke@nyulangone.org (J.D.B.)
†These authors contributed equally to this work.
Cite this article as S. Pinglay *et al.*, *Science* 377, eabk2820 (2022). DOI: 10.1126/science.abk2820

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abk2820

RESEARCH ARTICLE

GENE REGULATION

Synthetic regulatory reconstitution reveals principles of mammalian *Hox* cluster regulation

Sudarshan Pinglay[1]†, Milica Bulajić[2]†‡, Dylan P. Rahe[2], Emily Huang[1], Ran Brosh[1], Nicholas E. Mamrak[1], Benjamin R. King[1]§, Sergei German[1], John A. Cadley[1], Lila Rieber[3]¶, Nicole Easo[1], Timothée Lionnet[1,4,5], Shaun Mahony[3], Matthew T. Maurano[1,6], Liam J. Holt[1,5,7], Esteban O. Mazzoni[2]*, Jef D. Boeke[1,5,7]*

Precise *Hox* gene expression is crucial for embryonic patterning. Intra-*Hox* transcription factor binding and distal enhancer elements have emerged as the major regulatory modules controlling *Hox* gene expression. However, quantifying their relative contributions has remained elusive. Here, we introduce "synthetic regulatory reconstitution," a conceptual framework for studying gene regulation, and apply it to the *HoxA* cluster. We synthesized and delivered variant rat *HoxA* clusters (130 to 170 kilobases) to an ectopic location in the mouse genome. We found that a minimal *HoxA* cluster recapitulated correct patterns of chromatin remodeling and transcription in response to patterning signals, whereas the addition of distal enhancers was needed for full transcriptional output. Synthetic regulatory reconstitution could provide a generalizable strategy for deciphering the regulatory logic of gene expression in complex genomes.

Mammalian *Hox* genes encode deeply conserved transcription factors (TFs) that are organized into four dense clusters (*HoxA* to *HoxD*) lacking other coding genes (*1*). In response to developmental morphogens, *Hox* genes are expressed along the anterior-posterior axis of the developing embryo in a spatial and temporal pattern that mirrors the organization of the genes within the cluster (*2*–*4*). The alteration of this "colinear" *Hox* gene expression pattern results in gross developmental defects or diseases such as cancer (*5*, *6*).

In undifferentiated cells, where no *Hox* genes are expressed, the entire *HoxA* cluster (~130 kb) is targeted by Polycomb repressive complex 2 (PRC2) and marked by heterochromatin (*7*). Patterning signals such as retinoic acid (RA) and Wnt activate transcription through their downstream TFs. In response to RA, activated RA receptors (RARs) are bound to retinoic acid response elements (RAREs) found within the *HoxA* cluster. This is correlated with the separation of the cluster into two domains that contain transcribed and repressed genes, respectively (*8*–*12*) (Fig. 1A). RARE mutations can lead to decreased or abolished expression of the neighboring *HoxA* genes in the central nervous system, solidifying the notion that RAREs directly control *Hox* gene expression (*9*, *13*, *14*). However, the relationship between RARE activity and chromatin domain formation is not yet well established.

A collection of intricate genetic manipulations has revealed a complex regulatory landscape surrounding the *Hox* clusters (*15*–*19*). For example, the *HoxA* cluster relies on several RA and Wnt responsive distal regulatory elements (enhancers) located in the gene-poor region between *Hoxa1* and the next gene, *Skap2* (~300 kb away). Deleting some of these enhancers reduces, but never eliminates, *HoxA* expression in response to RA or Wnt (*20*–*22*). Distal enhancer access to the genes in the inactive domain is restricted by the formation of a strong topological boundary at CCCTC-binding factor (CTCF) sites within the cluster upon differentiation. Disruption of this boundary results in the misexpression of posterior *HoxA* genes in response to RA (*15*, *23*, *24*).

Thus, different regulatory modules are integrated to control *Hox* gene expression: local TF binding, distal enhancers, and topological DNA organization. However, a synergistic model describing their relative contribution and interactions has remained elusive (Fig. 1B).

The relative contribution of a cis-regulatory module could be measured by (i) generating a variant allele that only contains the elements that constitute the module, and (ii) isolating the variant allele from confounding factors, such as the compensatory effect of other regulatory elements in cis. Testing the necessity of individual elements by generating loss-of-function variants has received a lot of attention. However, tests of sufficiency have remained intractable owing to the inability to precisely manipulate DNA at a scale that accurately models the complexity and size of native loci (>100 kb) (*25*).

Short reporter constructs enable the study of variants only over a small genomic window (<10 kb) and also suffer from a lack of controlled genomic context, as they are largely randomly integrated or reside on episomal plasmids (*26*). Although they can test the ability of a sequence to drive expression, they lack the scale to study *Hox* cluster chromatin domain formation (>100 kb). Larger bacterial or yeast artificial chromosomes (BACs/YACs) can provide more genomic context and have been elegantly employed to understand regulation at *Hox* clusters (*27*–*29*). However, they are not easy to manipulate, making it challenging to generate a large number of variants that can be tested in vivo (*26*, *30*, *31*). Further, these large constructs are often randomly integrated into the genome, which confounds comparison across different constructs as a result of position effects and the integration of multiple copies. Methods for precise, single-copy integration of large DNA molecules in mammalian cells have not yet been applied to the study of *Hox* cluster regulation (*32*, *33*). Despite the revolution in genome editing, it is still inefficient and time-consuming to make intricate structural rearrangements or multiple defined edits in cis, phased on a single homolog (*34*, *35*).

We recently described a pipeline that harnesses the endogenous homologous recombination machinery in yeast to de novo assemble ~100-kb regions of mammalian genomes and integrate them into a defined location in mouse embryonic stem cells (mESCs) (*36*). This bottom-up assembly of genomic loci enables the introduction of an arbitrary number of variants in cis that are independent of any natural template.

Here, we apply and extend this technology to study the relative contributions of genomic context, distal enhancers, and intracluster regulatory elements to *HoxA* regulation. We constructed variants of the *HoxA* cluster (ranging from 130 to 170 kb) encoding combinations of the previously identified regulatory modules and integrated them into an ectopic locus, thereby isolating them from the native genomic neighborhood. We then asked whether the variant ectopic clusters were sufficient to reconstitute the transcriptional and epigenetic *HoxA* response to activating patterning signals (Fig. 1C).

Synthetic *HoxA* strategy and construction

All *HoxA* constructs described here were derived from rat (*Rattus norvegicus*) *HoxA* sequence, which shares ~90% sequence identity

[1]Institute for Systems Genetics, NYU Langone Health, New York, NY 10016, USA. [2]Department of Biology, New York University, New York, NY 10003, USA. [3]Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA. [4]Department of Cell Biology, NYU Langone Health, New York, NY 10016, USA. [5]Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY 11201, USA. [6]Department of Pathology, NYU Langone Health, New York, NY 10016, USA. [7]Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA.
***Corresponding author. Email: eom204@nyu.edu (E.O.M.); jef.boeke@nyulangone.org (J.D.B.)**
†These authors contributed equally to this work.
‡Present address: Biogen Inc., Cambridge, MA 02142, USA.
§Present address: Bristol-Myers Squibb Co., Lawrenceville, NJ 08648, USA.
¶Present address: Element Genomics Inc., Durham, NC 27701, USA.

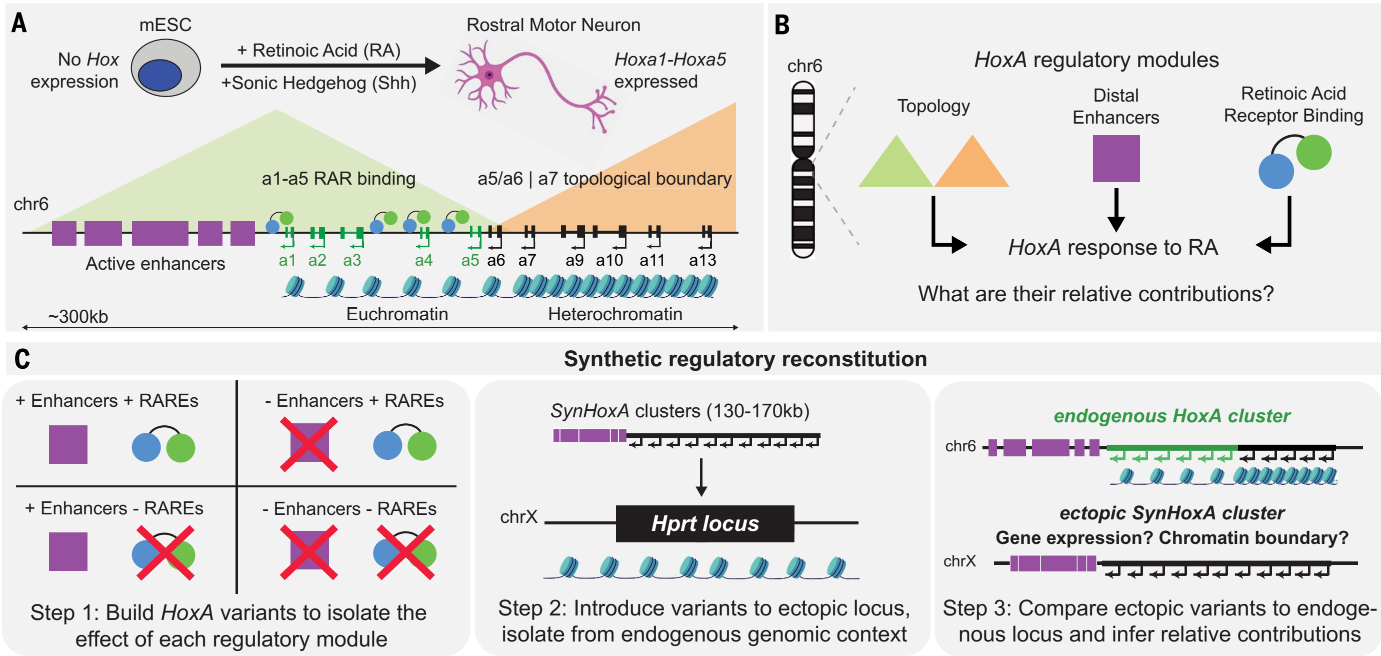


Fig. 1. *HoxA* regulation relies on the integration of multiple regulatory modules. (**A**) Schematic of *HoxA* regulation in response to retinoic acid (RA) during in vitro mouse ES cell (mESC)–motor neuron differentiation. (**B**) Multiple regulatory modules, including enhancers, TF binding, and topology, are integrated to drive *HoxA* cluster response to RA. (**C**) Schematic of the synthetic regulatory reconstitution approach. Synthetic *HoxA* variants encoding various combinations of regulatory modules are built and integrated at an ectopic location in the genome. The response of these synthetic ectopic clusters to RA reveals their sufficiency and relative contribution in driving faithful *HoxA* expression.

with the mouse *HoxA* sequence. This facilitated experiments in two distinct genetic backgrounds: (i) cells containing the endogenous *HoxA* cluster ($HoxA^{+/+}$), allowing direct comparisons of expression levels to an internal control, and (ii) cells lacking endogenous mouse *HoxA* to eliminate potential sequence-mapping challenges ($HoxA^{-/-}$).

We first constructed a 134-kb wild-type minimal rat *HoxA* cluster (*SynHoxA*) (Fig. 2 and fig. S1). *SynHoxA* contains all *HoxA* coding genes and encompasses the sequence corresponding to the contiguous domain repressed by H3K27me3 (trimethylated histone H3 Lys^{27}) in undifferentiated mESCs. We produced polymerase chain reaction (PCR) amplicons from BACs bearing the rat *HoxA* cluster, with overlap between adjacent segments to enable homologous recombination. These amplicons were recombined in yeast to produce the 134-kb *SynHoxA* construct, termed an "assemblon" (fig. S1B). Edits to the assemblon can be made by switching the wild-type amplicons with synthetic DNA bearing the desired changes or by editing the assemblons directly using highly efficient, marker-free CRISPR/Cas9-based engineering in yeast (*37*). Assemblons were recovered from yeast into bacteria to purify large amounts of DNA (Fig. 2A). To test the contribution of distal enhancers to *HoxA* regulation, we fused all experimentally verified distal regulatory elements directly upstream of the core *SynHoxA* assemblon, generating the 170-kb *Enhancers+SynHoxA* construct (fig. S2).

DNA sequencing at each step confirmed assemblon integrity (Fig. 2C and fig. S3). We detected mutations at a frequency of ~1 nucleotide per 6 kb, likely arising from PCR errors (fig. S3). We verified that none of these mutations were likely to affect the function of the clusters in our differentiation system described below (table S1).

All constructs were delivered to the *Hprt* locus on the X chromosome of both $HoxA^{+/+}$ and $HoxA^{-/-}$ mESCs using Cre-mediated recombination (*38*) (figs. S4 and S5). We reasoned that *Hprt* would be an appropriate site to attempt regulatory reconstitution because it is a housekeeping gene with little regulatory activity in the surrounding regions, and its use as a "safe harbor" site is well documented (*39*, *40*). Furthermore, transposing the *HoxA* cluster to an open locus would be the most stringent test of its ability to recruit the repressive PRC2 complex in undifferentiated cells. We used capture sequencing to verify that mESC clones contained the entire assemblon, specifically at Hprt and in single copy (*41*) (Fig. 2B).

Induction of *SynHoxA* gene expression during motor neuron differentiation

We investigated the response of the ectopic clusters to RA in a widely used differentiation system that recapitulates key aspects of ventral spinal cord development, including the activation of the appropriate *HoxA* genes (fig. S6A) (*42*, *43*). Endogenous *HoxA* genes are repressed before exposure to patterning signals, and *Hoxa1–5* are induced in response to the "anterior" RA signal (*10*).

We performed RNA-seq on wild-type $HoxA^{+/+}$ cells containing ectopic *SynHoxA* variants over the course of the differentiation protocol. Comparison to previously published control datasets by principal components analysis (PCA) revealed that the samples grouped largely according to time since RA treatment (*44*, *45*) (fig. S6B). Cells containing *SynHoxA* variants down-regulated the expression of pluripotency markers, up-regulated the expression of markers of motor neuron differentiation, and induced *Hoxa1–5* from the endogenous cluster upon exposure to RA as expected (Fig. 3, A and C, and fig. S6, C and D). Thus, the presence of an ectopic synthetic cluster does not affect the ability of cells to differentiate appropriately after RA treatment.

The ectopic *Enhancers+SynHoxA* cluster induced *SynHoxa1–5* starting 24 hours after RA treatment (Fig. 3B and figs. S7 and S8, A and B), whereas *SynHoxa6* to *SynHoxa13* remained repressed throughout. Similarly, ectopic *SynHoxA* specifically induced *SynHoxa1–5* (Fig. 3D and figs. S7 and S8C) without misexpression of any posterior genes. Thus, neither endogenous genomic context nor the wide spacing of enhancer elements is strictly required

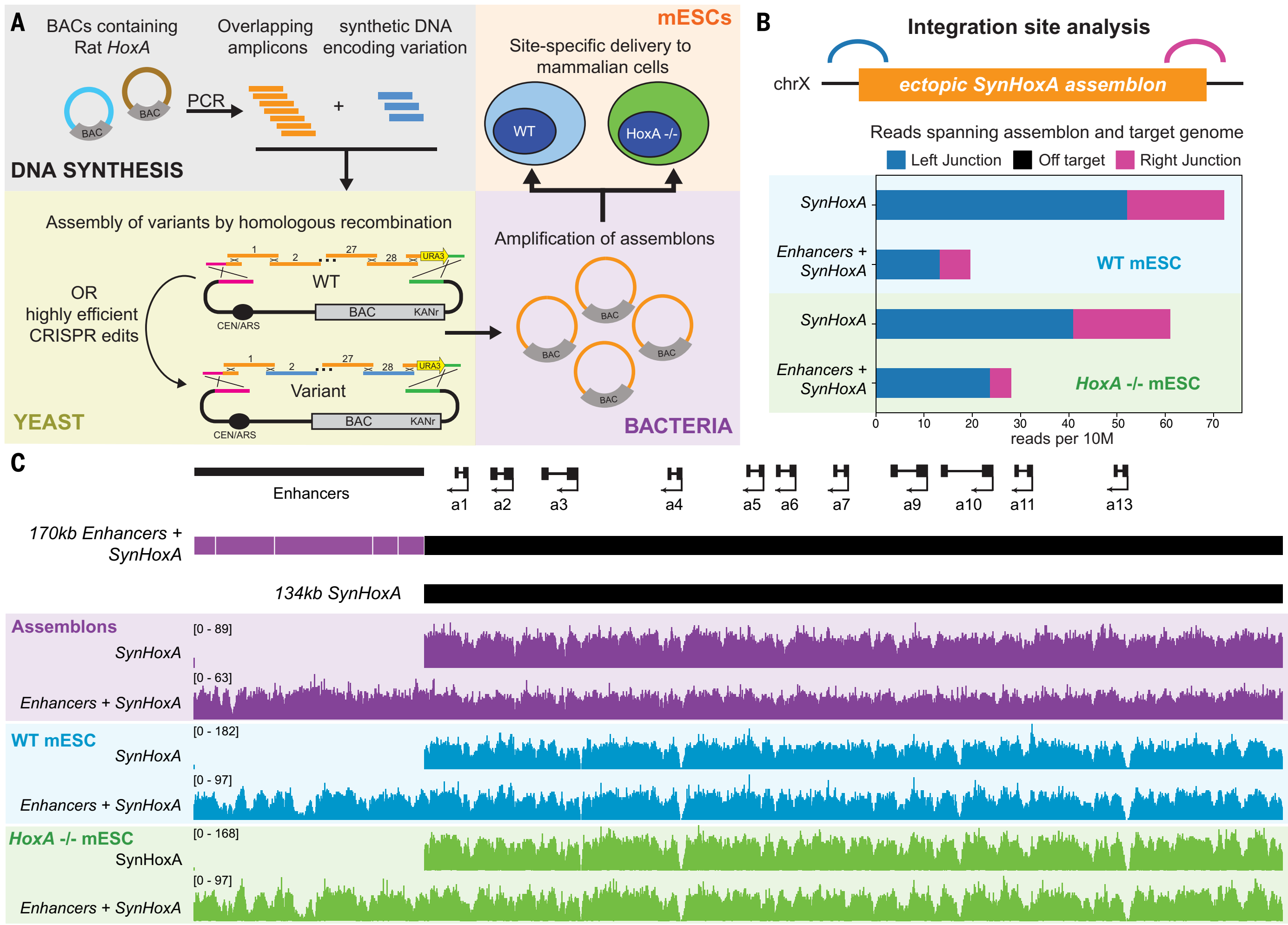


Fig. 2. Build and delivery of *SynHoxA* constructs. (**A**) Schematic of the process to generate mESCs bearing ectopic synthetic *HoxA* clusters via homologous recombination-based assembly in yeast and amplification in bacteria. (**B**) Analysis of reads spanning the synthetic construct and the host genome by bamintersect revealed only the expected junctions with no off-target integrations. (**C**) Schematic of the 134-kb *SynHoxA* and 170-kb *Enhancers+SynHoxA* constructs. Sequencing data for assemblon DNA isolated from bacteria (purple) and from capture sequencing after integration into wild-type (blue) and $HoxA^{-/-}$ (green) mESCs.

for a *Hox* cluster to induce the appropriate genes.

We quantified gene induction from the ectopic clusters by comparing each gene to its endogenous mouse *HoxA* counterpart (Fig. 3, E and F). Both ectopic clusters induced lower *SynHoxa1* transcription, whereas *SynHoxa2* induction surpassed the endogenous gene. The induction kinetics of *SynHoxa3–5* were slower than *Hoxa3–5*, but the mRNA levels became comparable after 96 hours, particularly in the *Enhancers+SynHoxA* construct. In general, *Enhancers+SynHoxA* induced higher levels of *SynHoxa1–5* expression than *SynHoxA*, especially at early time points (Fig. 3 and figs. S7 and S8, B and C).

In our differentiation model, temporal colinearity of genes from the endogenous *HoxA* cluster is exhibited by (i) an early induction of *Hoxa1* (peaks at 24 hours) and its subsequent down-regulation and (ii) the sequential induction of *Hoxa2–5* from 24 to 96 hours. At 96 hours, the expression level is *Hoxa2* < *Hoxa3* < *Hoxa4* < *Hoxa5* (figs. S7 to S9). The minimal *SynHoxA* cluster does not recapitulate the temporal expression pattern of *Hoxa1*. *SynHoxa1* is expressed at low levels throughout but increases steadily with time, mirroring the regulation of *SynHoxa2–5* genes. However, this is partially rescued with the addition of enhancers in *Enhancers+SynHoxA*. *SynHoxa1* expression in this context peaks at 48 hours and decreases at 96 hours. The temporal expression of *SynHoxa2–5* is retained in both constructs (fig. S9).

The observed bulk differences in gene expression between the ectopic clusters could be attributed either to a change in the number of cells that induce expression or to expression level changes in each cell. To address this, we investigated the response of the synthetic clusters to RA at the single-cell level using (i) Hoxa5 antibody staining and (ii) RNA single-molecule fluorescence in situ hybridization for anterior *SynHoxA* genes. We performed these experiments in the $HoxA^{-/-}$ background. Gene expression from the ectopic clusters was not influenced by the presence or absence of the endogenous *HoxA* cluster (fig. S10), and the cells differentiated appropriately (fig. S11). In both methods, the observed bulk differences in expression were not due to a difference in the number of cells that express anterior *SynHoxA* genes, but due to an increase in the amount of expression per cell in the presence of the enhancers (figs. S12 to S15).

We investigated whether these findings could be extended to other patterning signals beyond RA, such as the posteriorizing Wnt signal (*10*, *22*). Both ectopic clusters induced more posterior genes upon treatment with Wnt, up to *SynHoxa11* (fig. S16). Intriguingly, the addition of distal enhancers does not seem

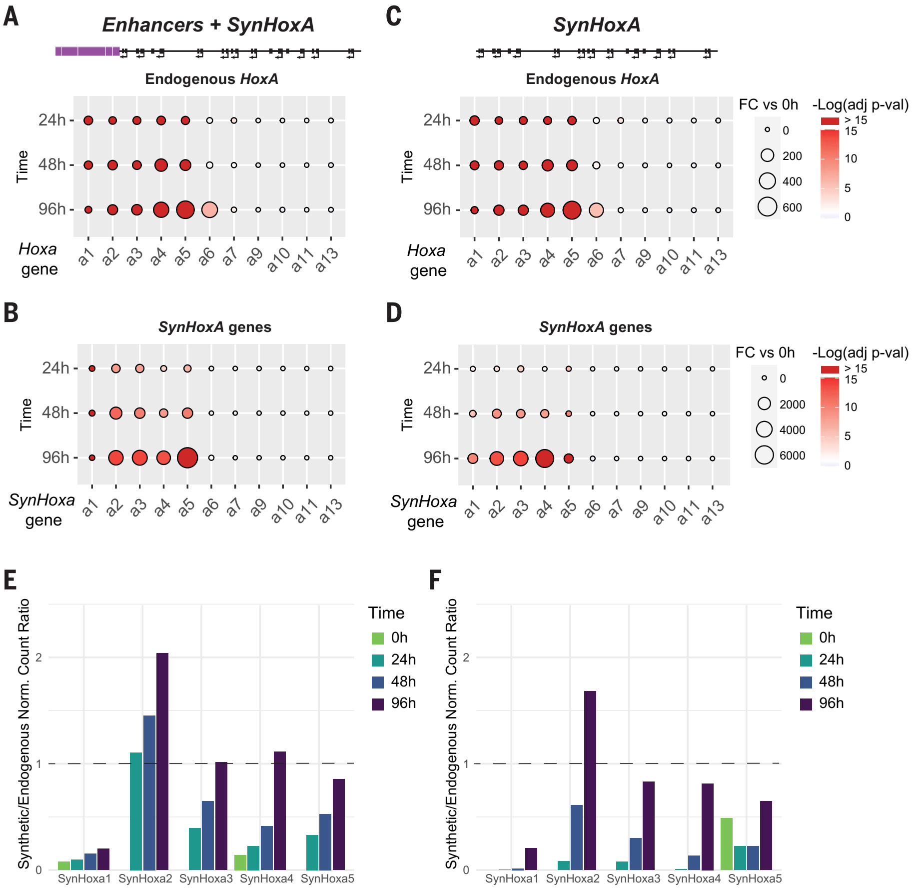


Fig. 3. *SynHoxA* variants up-regulate the correct subset of genes in response to RA patterning signal. (**A** to **D**) Fold change (FC) of RNA-seq data for endogenous mouse *HoxA* [(A) and (C)] and *SynHoxA* [(B) and (D)] genes during RA differentiation, relative to expression before RA treatment (n = 2). *SynHoxA* variants induce the correct genes (*SynHoxa1–5*) in response to RA. (**E** and **F**) Ratios of gene expression for *SynHoxA* genes to endogenous mouse *HoxA* genes (n = 2). Counts for the endogenous *HoxA* genes were halved to normalize for two endogenous *HoxA* versus one ectopic *SynHoxA* copy.

to consistently modify the Wnt response across *SynHoxa* genes as much as it did for RA.

Together, these data support a model in which the *HoxA* cluster contains all the necessary information to decode patterning signals into the appropriate positional identity, independent of distal enhancers and the native genomic architecture. Distal enhancers are not required to specify active genes but have a critical role in modulating transcriptional output in response to RA. Whereas both ectopic clusters induce the correct subset of genes, additional elements may be required to fine-tune expression amplitude and timing.

Chromatin reorganization at *SynHoxA* clusters through differentiation

In undifferentiated cells, *Hox* clusters are carpeted with repressive H3K27me3 marks and recruit CTCF to potential boundary positions established in response to extracellular signals. The endogenous *HoxA* cluster responds to RA by separating into an active and inactive domain, forming a precise chromatin boundary between *Hoxa5* and *Hoxa6* (*10*, *23*). We investigated the chromatin dynamics and CTCF recruitment of the relocated *HoxA* clusters by performing chromatin immunoprecipitation sequencing (ChIP-seq) in the $HoxA^{-/-}$ background to more reliably map reads to the ectopic clusters, thereby enhancing resolution.

Prior to differentiation, both *Enhancers+SynHoxA* and *SynHoxA* were entirely covered with high levels of H3K27me3. In addition, CTCF was recruited to the appropriate sites (Fig. 4, A and B). Thus, the ability to recruit all components for correct patterning is intrinsic to the *HoxA* cluster sequence and is independent of endogenous genomic context or distal enhancers. Upon RA treatment, *Enhancers+SynHoxA* remodeled chromatin in a manner that is similar to the endogenous cluster. H3K27me3 decreased in the *SynHoxa1–5* domain and increased at *SynHoxa6–13* (Fig. 4C). An increase in acetylated H3K27 (H3K27ac) complemented H3K27me3 removal from *SynHoxa1-5*. H3K27me3 was entirely cleared by 48 hours, slightly slower than the endogenous locus, which clears by 24 hours (*10*). Thus, the endogenous genomic context at *HoxA* is not required to translate an extracellular signal into an accurate epigenetic chromatin state.

The minimal *SynHoxA* cluster also recruited H3K27ac to anterior genes upon RA activation (Fig. 4D). Unlike the endogenous *HoxA* cluster and *Enhancers+SynHoxA*, H3K27me3 was not entirely lost from the *SynHoxa1-5* domain during differentiation (Fig. 4D). Nonetheless, *SynHoxA* formed the appropriate, albeit weak, chromatin boundary at the *SynHoxa5–a6* CTCF binding site (Fig. 4E and fig. S17, A and B). Thus, the minimal *SynHoxA* cluster has the intrinsic ability to induce dynamic chromatin domains, independent of genomic context or enhancers.

Topological organization of ectopic *SynHoxA* clusters

The three-dimensional structure of the *HoxA* cluster changes during motor neuron differentiation, transitioning from a single association domain to two domains containing active or repressed chromatin (*15*, *23*, *24*). We investigated the topological organization of the ectopic clusters by performing Hi-C, a technique for detecting genome-wide chromatin interactions, during differentiation (0 and 48 hours after RA treatment). Both clusters formed self-associating domains in undifferentiated cells without generating a de novo topologically associating domain (TAD) boundary (fig. S18). Similar to the endogenous cluster, *Enhancers+SynHoxA* broke into two domains during differentiation, with enhancers and active genes in one domain and repressed genes in the other (fig. S18A). The minimal *SynHoxA* similarly transitioned from a compact self-associated state in undifferentiated cells into two domains during differentiation (fig. S18B). We did not observe strong evidence for trans-chromosomal interactions between the ectopic locus and its endogenous enhancers (tables S2 and S3). Thus, the ectopic *HoxA* clusters have the intrinsic ability to self-organize in three dimensions, mirroring the expression and chromatin changes that occur upon differentiation.

The RARE sites within the *HoxA* cluster are required for the RA response

The minimal *SynHoxA* transformed the RA signal into the correct transcriptional and chromatin programs. This behavior could theoretically depend on RAREs located within the *Hoxa1–5* domain, other sequences within the cluster, or interactions with other regulatory elements at the ectopic locus. To distinguish between these possibilities, we built a third construct lacking both RAREs and enhancers (*RAREΔ SynHoxA*) and integrated it into wild-type and $HoxA^{-/-}$ mESCs (figs. S19 and S20) (*9*).

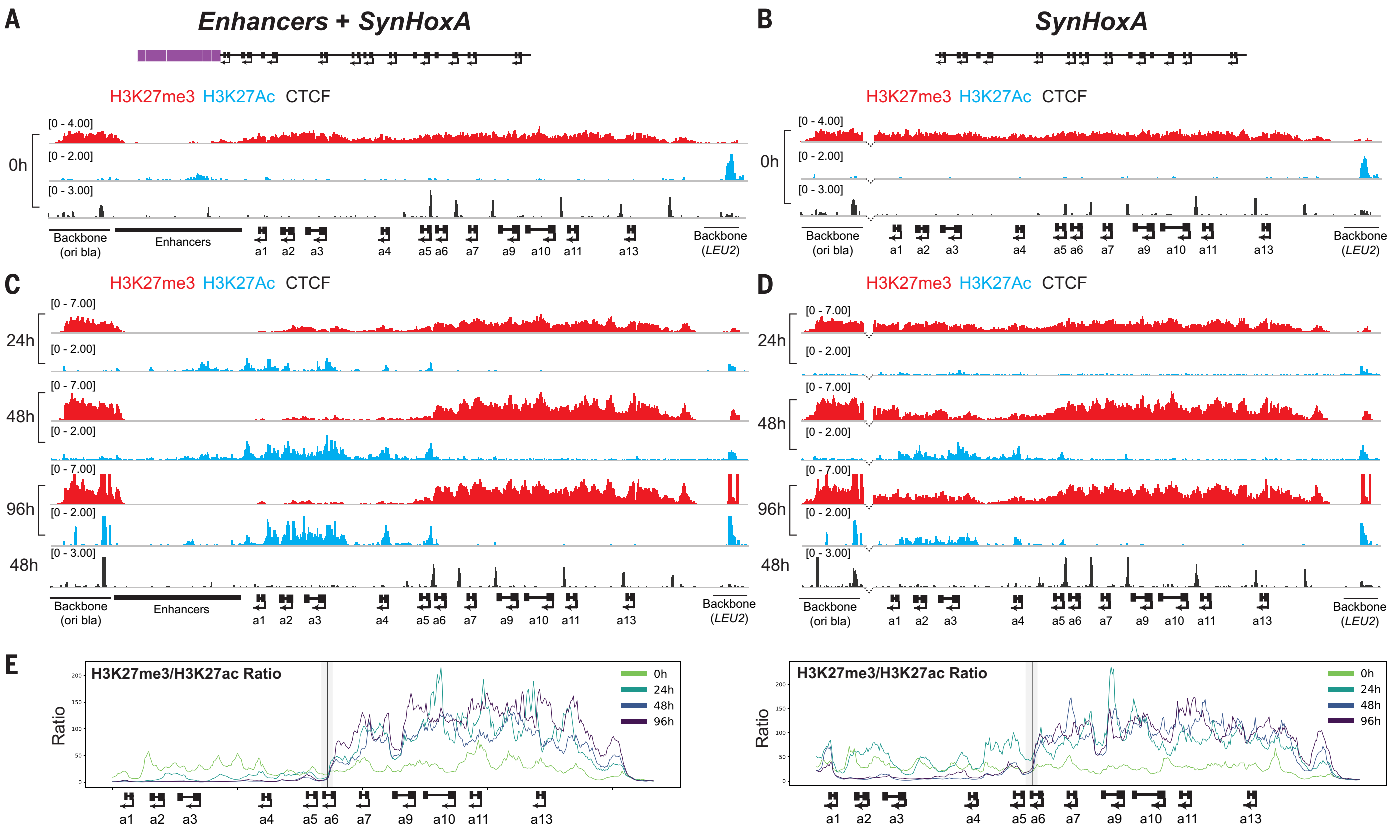


Fig. 4. Distal enhancers are required for full clearance of repressive chromatin and the formation of a sharp chromatin boundary. ChIP-seq analysis (n = 2) of activating H3K27ac (blue), repressive H3K27me3 (red), and CTCF (black) at *SynHoxA* clusters. (**A** and **B**) In undifferentiated cells, *SynHoxA* clusters contained H3K27me3, lacked H3K27ac, and recruited CTCF to correct locations. (**C** and **D**) In response to RA, the *SynHoxa1–5* domain contained activating H3K27ac and cleared repressive H3K27me3. Dotted lines at the anterior breakpoint between the cluster and the vector sequences in (B) and (D) indicate presence of enhancer sequences in the reference genome to which no reads are mapped. (**E**) RPKM (reads per kilobase per million mapped reads) normalized ratios of repressive H3K27me3 to active H3K27ac across *SynHoxA*. The black line marks the *Hoxa5-Hoxa6* CTCF site; gray shading indicates the windows contributing to ChIP-seq signal at the site.

Cells carrying *RAREΔSynHoxA* differentiated and induced endogenous *HoxA* genes appropriately (Fig. 5A and fig. S20, D and E). The ectopic cluster was decorated with H3K27me3 in mESCs, indicating that RAREs are not required to recruit the repressive chromatin machinery. However, unlike the previous constructs, *RAREΔ SynHoxA* failed to up-regulate the expression of *SynHoxa1–5* or to form a chromatin boundary in response to RA (Fig. 5, B to D, and figs. S8D and S17C). *SynHoxa5* was the only gene with any signal, potentially due to a poorly characterized RAR binding site located between *Hoxa5* and *Hoxa6* (*9*, *10*). This null phenotype indicates that the behavior of ectopic *SynHoxA* clusters reflects their innate regulatory potential and not the effect of novel interactions formed at the *Hprt* locus.

This lack of RA response provided the ideal background to measure the independent contribution of distal enhancers to *HoxA* gene regulation. To that end, we built and integrated a fourth construct: *Enhancers+RAREΔ SynHoxA*, with the distal enhancers inserted upstream of *RAREΔ SynHoxA* (figs. S19 and S20). In *Enhancers+RAREΔ SynHoxA*, *SynHoxa2–5* were induced at low levels in response to RA (Fig. 6, A and B, and fig. S8E). A faint chromatin boundary formed at the appropriate location between *SynHoxa5* and *SynHoxa6* (Fig. 6, C and D, and fig. S17D). This suggests that the distal enhancers may have a weak ability to activate *HoxA* gene transcription without the driving force provided by internal RAREs, or that they synergize with the poorly characterized RARE still present in this construct.

Discussion

We developed a "synthetic regulatory reconstitution" approach to characterize the relative contributions of distal enhancers and intracluster TF binding in specifying distinct aspects of *HoxA* regulation. A minimal ectopic cluster lacking distal enhancers induced the correct subset of genes in response to two developmental signals (RA and Wnt). The presence of distal enhancers increased the transcriptional output from the ectopic cluster in response to RA, especially at earlier time points. These results are consistent with previous studies in which deletion of enhancers from the endogenous locus resulted in lower *Hox* gene transcription in response to morphogens (*20*, *22*, *46*). At the time point we investigated, the ectopic cluster with distal enhancers did not induce higher expression in response to Wnt. This may indicate that some Wnt-responsive enhancers are missing in these constructs, or that the distal enhancers are ineffective in this configuration. Alternatively, earlier time points may reveal differences in transcriptional dynamics or output between these constructs.

The importance of both RAREs and distal enhancers to *Hox* gene expression was firmly established before this work. However, it was difficult to quantify their relative contributions to the establishment of transcriptional and chromatin domains at the scale of an entire *Hox* cluster with previous techniques. Our results show that an ectopic cluster lacking all previously described internal RAREs (*RAREΔ SynHoxA*) failed to respond to RA both at the transcriptional and chromatin levels. This gene expression phenotype was not fully rescued by

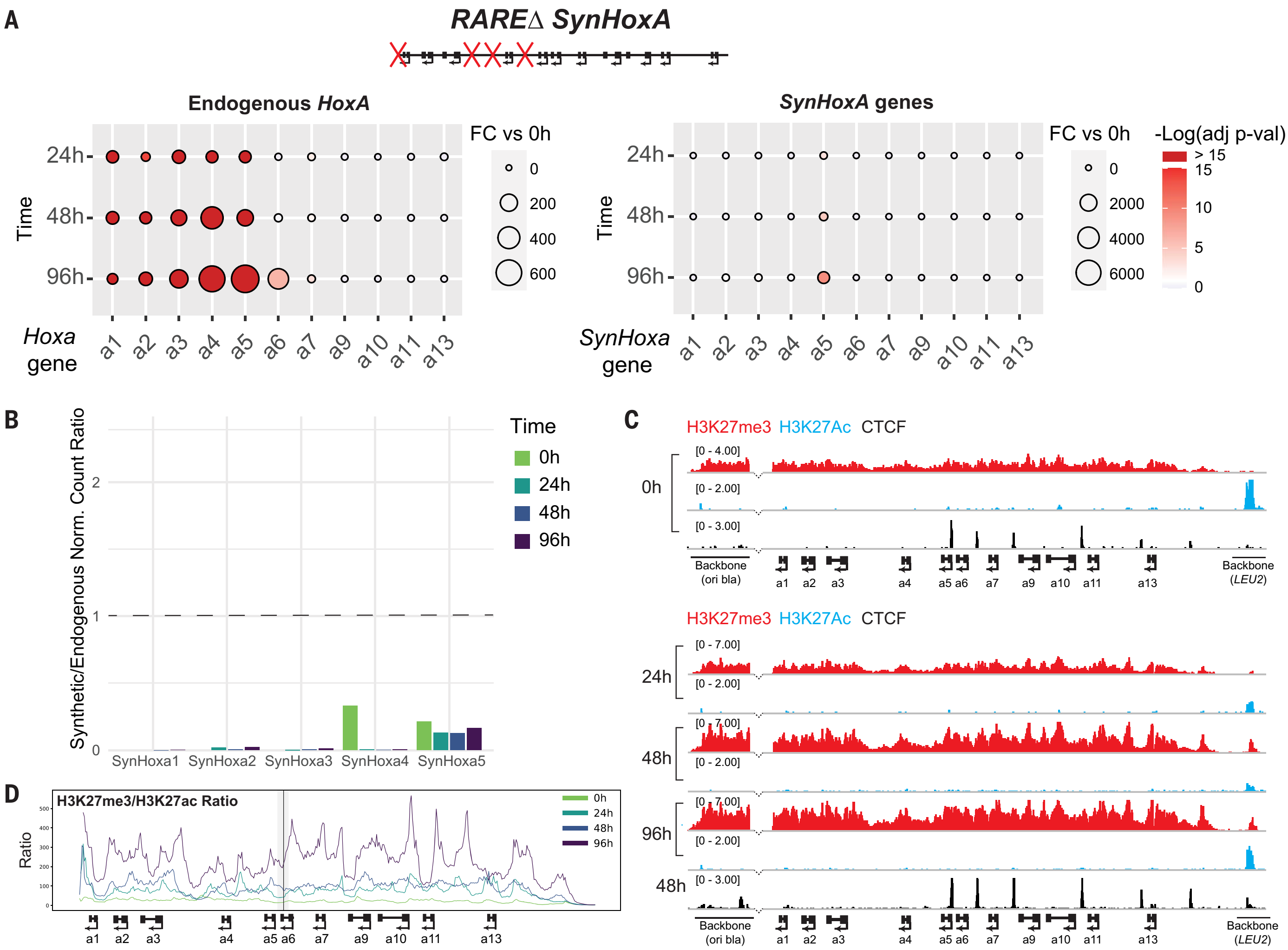


Fig. 5. Retinoic acid receptor response element (RARE) sites are required for the RA response. (**A**) Fold change of RNA-seq data for endogenous *HoxA* and *SynHoxA* genes during differentiation (*n* = 2). (**B**) Ratios of gene expression for *SynHoxA* genes to endogenous mouse *HoxA* genes (*n* = 2). (**C**) ChIP-seq revealed no evidence of H3K27me3 (red) clearance and H3K27ac (blue) recruitment at *RARE*Δ *SynHoxA*. Dotted lines show the anterior breakpoint between the cluster and the vector sequences in (C) as in Fig. 4. (**D**) Ratio of repressive H3K27me3 to active H3K27ac chromatin across *SynHoxA* as in Fig. 4E.

the addition of distal enhancers in *Enhancers +RARE*Δ *SynHoxA*. However, some chromatin remodeling was observed. Together, removal of internal RAREs led to virtually complete loss of gene expression, whereas removal of enhancers led to a reduction in expression at early time points, with almost complete rescue at later time points. Therefore, distal enhancers do not induce high levels of transcription or induce drastic chromatin remodeling in the absence of intracluster RAREs, but synergize with the RAREs to play a critical role in fine-tuning expression levels (Fig. 6, E and F).

All ectopic clusters recruited CTCF and PRC2 in embryonic stem cells, implying that this property is *Hox* cluster–intrinsic. Therefore, precise CTCF positioning within the cluster does not depend on interactions with other elements at the endogenous TAD boundary. Together, our results imply that *Hox* clusters are discrete units with an intrinsic ability to respond to patterning signals, strengthening original observations made at the *HoxD* cluster using random BAC transgenesis (*28*, *29*). This is also congruent with the idea that the evolution of novel distal enhancers is a source of morphological novelty in secondary structures such as limbs (*16*, *47*).

The inability of the minimal *SynHoxA* to fully clear repressive marks in the anterior domain could have several causes. First, the boost in transcription provided by distal enhancers at early time points might facilitate the clearance of repressive chromatin. Second, the enhancers could serve as platforms to recruit additional chromatin modifiers. Finally, the vector backbone that is introduced as part of the delivery harbors repressive chromatin modifications throughout differentiation (Fig. 4). The spread of repressive chromatin from this region might be more effectively prevented by enhancer sequences that contain fewer CpG islands than the cluster itself. Future experiments using scarless delivery methods will enable us to distinguish between these hypotheses (*41*). Although we see no strong genetic or topological evidence for trans-chromosomal interactions, we cannot fully exclude the possibility that they may play some part in activating *SynHoxA* genes. Higher-resolution chromatin conformation data centered on the ectopic clusters may help to address this question.

This study represents a proof of principle for "synthetic regulatory reconstitution." Targeting large, fully editable constructs to precise genomic locations enables quantitative comparisons between variants and promises to address critical questions in gene regulation and genome organization. Multiple elements required for the finer analysis of constructs through

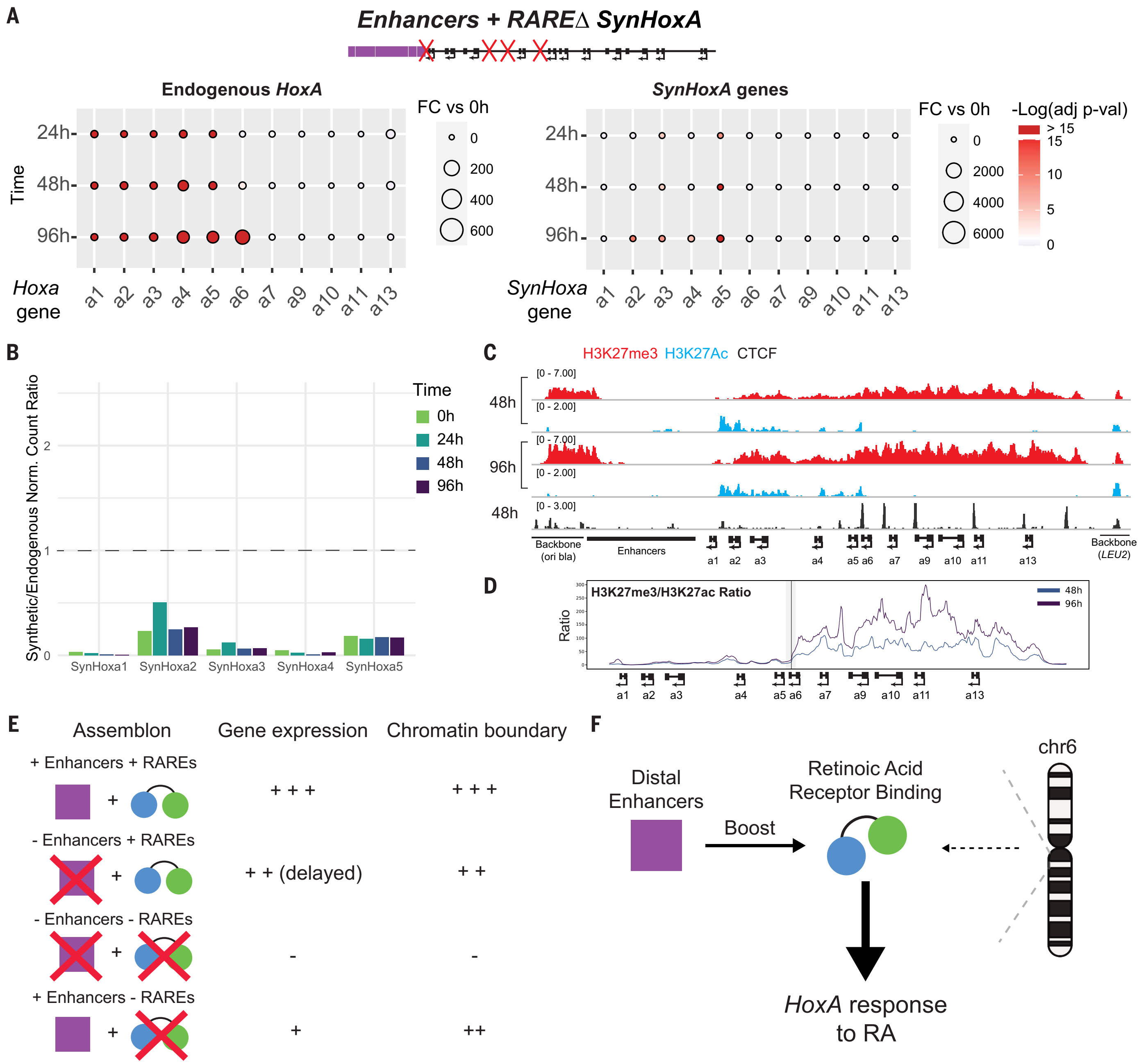


Fig. 6. The addition of enhancers to *SynHoxA RARE*Δ does not rescue gene expression. (**A**) Fold change of RNA-seq data for endogenous *HoxA* and *SynHoxA* genes during differentiation (n = 2). (**B**) Ratios of gene expression for *SynHoxA* genes to endogenous mouse *HoxA* genes (n = 2). (**C**) ChIP-seq revealed the appropriate recruitment of CTCF (black) and the formation of a weak chromatin boundary at *Enhancers+RARE*Δ *SynHoxA* upon differentiation (n = 2). (**D**) Ratio of repressive H3K27me3 to active H3K27ac chromatin across *SynHoxA* as in Fig. 4E. (**E**) Summary of gene expression and chromatin boundary phenotypes across all *SynHoxA* clusters. (**F**) Model describing relative contributions of distal enhancers, intra-*Hox* binding, and genomic context to the RA response at *HoxA*.

differentiation, such as live-cell imaging of transcription and chromatin mobility, could also be included via bottom-up synthesis (*48*). Testing different ectopic sites such as those marked with constitutive heterochromatin, cross-species transplants of regulatory landscapes, and phenotyping in richer systems such as living mice and gastruloids are all attractive avenues to explore (*49*).

Differences from endogenous gene expression dynamics were observed, even for the *Enhancers+SynHoxA* construct. Thus, even the largest construct does not contain all the regulatory information required for refining gene expression. The great value in pursuing the synthetic regulatory reconstitution strategy is realized in cases where endogenous regulation cannot be fully recapitulated. This points to gaps in knowledge that we can attempt to fill by building successively larger or more intricate ectopic constructs in the future until no differences are observed when compared to the endogenous cluster. Reconstitution is a powerful framework for dissecting complex biochemical processes because it allows for exquisite control over components of the system under study (*50*, *51*). By analogy,

our approach allows for the generation of locus-scale variant constructs with any combination of desired changes. We expect synthetic regulatory reconstitution to be a fundamental component of the toolbox for studying transcriptional regulation.

Methods summary

A full description of the methods can be found in the supplementary materials. In brief, *SynHoxA* constructs were fabricated in yeast and integrated into mESCs as described (*36*, *41*). *SynHoxA* mESCs were differentiated to motor neurons and characterized by RNA-seq, ChIP-seq, and Hi-C as described (*10*, *44*, *45*). Sequencing data were analyzed using custom pipelines.

REFERENCES AND NOTES

1. D. Duboule, The rise and fall of Hox gene clusters. *Development* **134**, 2549–2560 (2007). doi: 10.1242/dev.001065; pmid: 17553908
2. M. Kmita, D. Duboule, Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**, 331–333 (2003). doi: 10.1126/science.1085753; pmid: 12869751
3. W. McGinnis, R. Krumlauf, Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992). doi: 10.1016/0092-8674(92)90471-N; pmid: 1346368
4. D. Duboule, G. Morata, Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* **10**, 358–364 (1994). doi: 10.1016/0168-9525(94)90132-5; pmid: 7985240
5. E. B. Lewis, A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570 (1978). doi: 10.1038/276565a0; pmid: 103000
6. N. Shah, S. Sukumar, The Hox genes and their roles in oncogenesis. *Nat. Rev. Cancer* **10**, 361–371 (2010). doi: 10.1038/nrc2826; pmid: 20357775
7. R. Margueron, D. Reinberg, The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011). doi: 10.1038/nature09784; pmid: 21248841
8. S. Mahony *et al*., Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol.* **12**, R2 (2011). doi: 10.1186/gb-2011-12-1-r2; pmid: 21232103
9. C. Nolte, B. De Kumar, R. Krumlauf, Hox genes: Downstream "effectors" of retinoic acid signaling in vertebrate embryogenesis. *Genesis* **57**, e23306 (2019). doi: 10.1002/dvg.23306; pmid: 31111645
10. E. O. Mazzoni *et al*., Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat. Neurosci.* **16**, 1191–1198 (2013). doi: 10.1038/nn.3490; pmid: 23955559
11. D. Noordermeer *et al*., The dynamic architecture of Hox gene clusters. *Science* **334**, 222–225 (2011). doi: 10.1126/science.1207194; pmid: 21998387
12. N. Soshnikova, D. Duboule, Epigenetic temporal control of mouse Hox genes in vivo. *Science* **324**, 1320–1323 (2009). doi: 10.1126/science.1171468; pmid: 19498168
13. V. Dupé *et al*., In vivo functional analysis of the Hoxa-1 3′ retinoic acid response element (3′RARE). *Development* **124**, 399–410 (1997). doi: 10.1242/dev.124.2.399; pmid: 9053316
14. M. Frasch, X. Chen, T. Lufkin, Evolutionary-conserved enhancers direct region-specific expression of the murine Hoxa-1 and Hoxa-2 loci in both mice and Drosophila. *Development* **121**, 957–974 (1995). doi: 10.1242/dev.121.4.957; pmid: 7743939
15. N. Lonfat, T. Montavon, F. Darbellay, S. Gitto, D. Duboule, Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* **346**, 1004–1006 (2014). doi: 10.1126/science.1257493; pmid: 25414315
16. T. Montavon, D. Duboule, Chromatin organization and global regulation of Hox gene clusters. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120367 (2013). doi: 10.1098/rstb.2012.0367`pmid: 23650639
17. T. Montavon, N. Soshnikova, Hox gene regulation and timing in embryogenesis. *Semin. Cell Dev. Biol.* **34**, 76–84 (2014). doi: 10.1016/j.semcdb.2014.06.005; pmid: 24930771
18. T. Montavon *et al*., A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011). doi: 10.1016/j.cell.2011.10.023; pmid: 22118467
19. S. Berlivet *et al*., Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLOS Genet.* **9**, e1004018 (2013). doi: 10.1371/journal.pgen.1004018; pmid: 24385922
20. K. Cao *et al*., SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression. *Genes Dev.* **31**, 787–801 (2017). doi: 10.1101/gad.294744.116; pmid: 28487406
21. B. De Kumar *et al*., Analysis of dynamic changes in retinoid-induced transcription and epigenetic profiles of murine Hox clusters in ES cells. *Genome Res.* **25**, 1229–1243 (2015). doi: 10.1101/gr.184978.114; pmid: 26025802
22. R. Neijts *et al*., Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes Dev.* **30**, 1937–1942 (2016). doi: 10.1101/gad.285767.116; pmid: 27633012
23. V. Narendra *et al*., CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015). doi: 10.1126/science.1262088; pmid: 25722416
24. V. Narendra, M. Bulajić, J. Dekker, E. O. Mazzoni, D. Reinberg, CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016). doi: 10.1101/gad.288324.116; pmid: 28087711
25. N. Ostrov *et al*., Technological challenges and milestones for writing genomes. *Science* **366**, 310–312 (2019). doi: 10.1126/science.aay0339; pmid: 31624201
26. M. Gasperini, L. Starita, J. Shendure, The power of multiplexed functional analysis of genetic variants. *Nat. Protoc.* **11**, 1782–1787 (2016). doi: 10.1038/nprot.2016.135; pmid: 27583640
27. J. A. Lehoczky, J. W. Innis, BAC transgenic analysis reveals enhancers sufficient for Hoxa13 and neighborhood gene expression in mouse embryonic distal limbs and genital bud. *Evol. Dev.* **10**, 421–432 (2008). doi: 10.1111/j.1525-142X.2008.00253.x; pmid: 18638319
28. F. Spitz, F. Gonzalez, D. Duboule, A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405–417 (2003). doi: 10.1016/S0092-8674(03)00310-6; pmid: 12732147
29. F. Spitz *et al*., Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev.* **15**, 2209–2214 (2001). doi: 10.1101/gad.205701; pmid: 11544178
30. K. R. Peterson *et al*., Use of yeast artificial chromosomes (YACs) in studies of mammalian development: Production of beta-globin locus YAC mice carrying human globin developmental mutants. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5655–5659 (1995). doi: 10.1073/pnas.92.12.5655; pmid: 7539923
31. N. Heintz, BAC to the future: The use of bac transgenic mice for neuroscience research. *Nat. Rev. Neurosci.* **2**, 861–870 (2001). doi: 10.1038/35104049; pmid: 11733793
32. F. G. Liberante, T. Ellis, From kilobases to megabases: Design and delivery of large DNA constructs into mammalian genomes. *Curr. Opin. Syst. Biol.* **25**, 1–10 (2021). doi: 10.1016/j.coisb.2020.11.003
33. H. A. Wallace *et al*., Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128**, 197–209 (2007). doi: 10.1016/j.cell.2006.11.044; pmid: 17218265
34. N. S. McCarty, A. E. Graham, L. Studená, R. Ledesma-Amaro, Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* **11**, 1281 (2020). doi: 10.1038/s41467-020-15053-x; pmid: 32152313
35. K. Kraft *et al*., Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Rep.* **10**, 833–839 (2015). doi: 10.1016/j.celrep.2015.01.016; pmid: 25660031
36. L. A. Mitchell *et al*., De novo assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics* **218**, iyab038 (2021). doi: 10.1093/genetics/iyab038; pmid: 33742653
37. J. E. DiCarlo *et al*., Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013). doi: 10.1093/nar/gkt135; pmid: 23460208
38. M. Iacovino *et al*., Inducible cassette exchange: A rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* **29**, 1580–1588 (2011). doi: 10.1002/stem.715; pmid: 22039605
39. M. Jasin, M. E. Moynahan, C. Richardson, Targeted transgenesis. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8804–8808 (1996). doi: 10.1073/pnas.93.17.8804; pmid: 8799106
40. M. Gasperini *et al*., CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *Am. J. Hum. Genet.* **101**, 192–205 (2017). doi: 10.1016/j.ajhg.2017.06.010; pmid: 28712454
41. R. Brosh *et al*., A versatile platform for locus-scale genome rewriting and verification. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023952118 (2021). doi: 10.1073/pnas.2023952118; pmid: 33649239
42. H. Wichterle, I. Lieberam, J. A. Porter, T. M. Jessell, Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110**, 385–397 (2002). doi: 10.1016/S0092-8674(02)00835-8; pmid: 12176325
43. M. Peljto, H. Wichterle, Programming embryonic stem cells to neuronal subtypes. *Curr. Opin. Neurobiol.* **21**, 43–51 (2011). doi: 10.1016/j.conb.2010.09.012; pmid: 20970319
44. B. Aydin *et al*., Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes. *Nat. Neurosci.* **22**, 897–908 (2019). doi: 10.1038/s41593-019-0399-y; pmid: 31086315
45. M. Bulajić *et al*., Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns. *Development* **147**, dev.194761 (2020). doi: 10.1242/dev.194761; pmid: 33028607
46. G. Su *et al*., CTCF-binding element regulates ESC differentiation via orchestrating long-range chromatin interaction between enhancers and HoxA. *J. Biol. Chem.* **296**, 100413 (2021). doi: 10.1016/j.jbc.2021.100413; pmid: 33581110
47. R. Freitas, C. Gómez-Marín, J. M. Wilson, F. Casares, J. L. Gómez-Skarmeta, Hoxd13 contribution to the evolution of vertebrate appendages. *Dev. Cell* **23**, 1219–1229 (2012). doi: 10.1016/j.devcel.2012.10.015; pmid: 23237954
48. H. Sato, S. Das, R. H. Singer, M. Vera, Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatiotemporal Dynamics of Gene Expression. *Annu. Rev. Biochem.* **89**, 159–187 (2020). doi: 10.1146/annurev-biochem-011520-104955; pmid: 32176523
49. L. Beccari *et al*., Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018). doi: 10.1038/s41586-018-0578-0; pmid: 30283134
50. K. A. Ganzinger, P. Schwille, More from less - bottom-up reconstitution of cell biology. *J. Cell Sci.* **132**, jcs227488 (2019). doi: 10.1242/jcs.227488; pmid: 30718262
51. A. P. Liu, D. A. Fletcher, Biology under construction: In vitro reconstitution of cellular function. *Nat. Rev. Mol. Cell Biol.* **10**, 644–650 (2009). doi: 10.1038/nrm2746; pmid: 19672276

ACKNOWLEDGMENTS

We thank the Mazzoni, Boeke, and Holt labs as well as the Institute for Systems Genetics community for their support, M. Khalfan (Genomics Core Facility at NYU) for making the reform tool publicly available, B. Ragipani for preliminary analysis on motor neuron differentiation markers, N. Zesati and S. Arora for help with preliminary visualization of Hi-C data, the Experimental Pathology core at NYU Langone for help with sectioning, and J. Skok and D. Reinberg for their insights. **Funding:** Supported by NHGRI grant RM1HG009491 (J.B., M.T.M., and E.O.M.), NINDS grant R01NS100897 and NIGMS grant R01GM138876 (E.O.M.), New York State Stem Cell Science predoctoral training grant C322560GG (M.B.), NIH grants R01AG075272 and R01GM127538 and Melanoma Research Foundation Award 687306 (T.L.), and NIH grant F32CA239394 (B.R.K.). **Author contributions:** S.P., M.B., E.O.M., and J.D.B. conceived of the project. S.P., M.B., D.P.R., E.H., R.B., N.E.M., B.R.K., and N.E. performed experiments. S.P., M.B., D.P.R., L.R., N.E.M., S.M., and M.T.M. analyzed data. S.G. and J.A.C. provided computational support. T.L., M.T.M., L.J.H., E.O.M., and J.D.B. supervised research. S.P., M.B., E.O.M., and J.D.B. wrote the manuscript with input from all authors. **Competing interests:** J.D.B. is a founder and director of CDI Labs Inc.; a founder of and consultant to Neochromosome Inc.; a founder of, scientific advisory board member of, and consultant to ReOpen Diagnostics LLC; and past or present scientific advisory board member of Sangamo Inc., Modern Meadow Inc., Rome Therapeutics Inc., Sample6 Inc., Tessera Therapeutics Inc., and the Wyss Institute. The other authors declare no competing interests.

Data and materials availability: All sequencing data from this study are deposited in NCBI GEO under GSE190906. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk2820
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S17
References (*52*–*79*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 2 July 2021; accepted 27 May 2022
10.1126/science.abk2820

RESEARCH ARTICLE SUMMARY

BIOCHEMISTRY

Mutations linked to neurological disease enhance self-association of low-complexity protein sequences

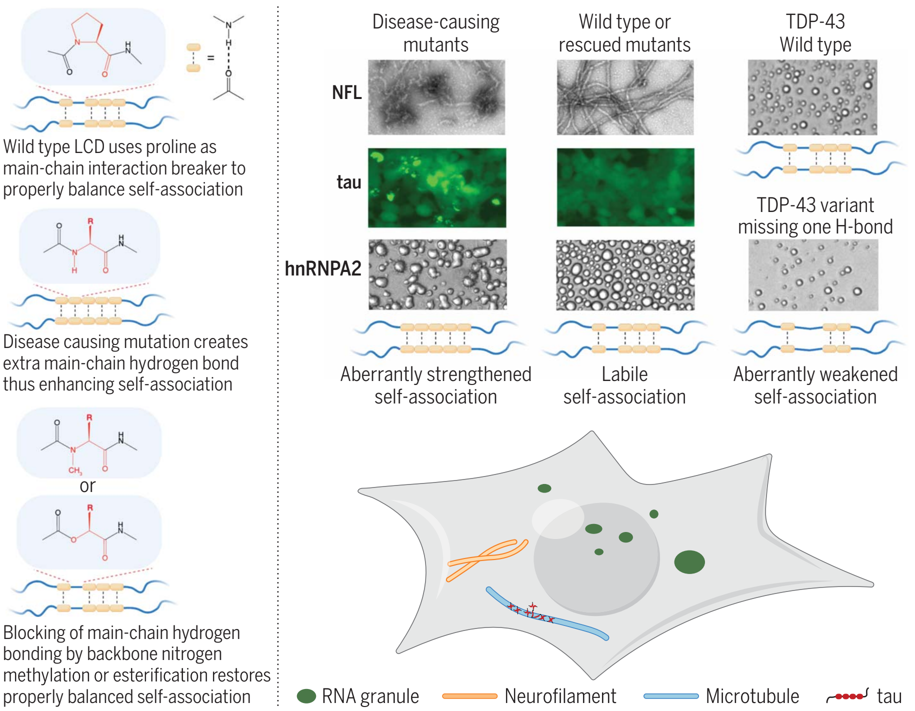
Xiaoming Zhou, Lily Sumrow, Kyuto Tashiro, Lillian Sutherland, Daifei Liu, Tian Qin, Masato Kato, Glen Liszczak*, Steven L. McKnight*

INTRODUCTION: Most proteins fold into stable, morphologically distinct three-dimensional shapes. By contrast, between 10 and 20% of the proteome in eukaryotic cells does not adopt stable structures. Protein domains that exhibit intrinsic disorder are typified by unusual sequences consisting of a limited representation of amino acid residues known as low-complexity domain (LCD) sequences. Protein LCDs are important to cell biology. They line the permeability channel of nuclear pores, allow for the assembly of intermediate filaments, and assist in the assembly of a variety of nuclear and cytoplasmic puncta that are not surrounded by lipid membranes. Despite their lack of folding, LCDs self-associate in a manner that allows phase separation from aqueous solution. Assays of phase separation have emerged as useful tools with which to study LCD self-association and its potential relationship to LCD function in living cells. Such studies have yielded conflicting observations. The prevailing thinking is that LCDs self-associate in the absence of any form of protein structure. An alternative perspective posits that association of LCDs is initiated by the formation of labile structures, and that this structural order is facilitated by hydrogen bond networks, allowing for favorable interchain adhesion in the form of cross-β interactions, as described 70 years ago by Linus Pauling.

RATIONALE: To test whether LCDs use main-chain hydrogen bonding to facilitate phase separation, variant LCDs associated with the RNA-binding protein TDP-43 RNA were constructed without individual hydrogen bonds across a region known to assemble in a cross-β conformation. These constructs allowed for experimental assessment of the importance of main-chain hydrogen bonding to phase separation. The role of main-chain hydrogen bonds was further investigated in studies of the neurofilament light (NFL) chain protein, the microtubule-associated tau protein, and the heterogeneous nuclear RNPA2 (hnRNPA2) RNA-binding protein to determine whether mutations known to cause human neurological disease might create an extra hydrogen bond, thereby enhancing the avidity of cross-β interactions within the LCDs of these proteins.

RESULTS: We prepared 23 variants of the TDP-43 LCD lacking the capacity for forming individual hydrogen bonds across a region critical to self-association and phase separation. Nine contiguous variants impeded phase separation. The locations of these variants corresponded precisely with the location of nearly identical regions of the cross-β structure reported independently by two different cryo–electron microscopy reconstructions of self-associated TDP-43 protein. Mutant variants that replace a proline to allow an extra hydrogen bond within LCDs of the disease-associated NFL, tau, and hnRNPA2 proteins were found to specify formation of inordinately stable self-associations. Aberrant stability was reversed by chemical modifications blocking the ability of the peptide nitrogen associated with the variant amino acid residues to participate in main-chain hydrogen bonding.

CONCLUSION: The LCDs that we studied function by forming labile structures in the form of specific cross-β assemblies. These structures can be weakened by the removal of single, main-chain hydrogen bonds or strengthened by human mutations specifying the addition of single hydrogen bonds, suggesting that they are poised at the threshold of thermodynamic equilibrium. The labile nature of these structures offers a dynamic dimension to many aspects of cell organization, as well as the opportunity for regulation by many forms of posttranslational modification. ■



The virtues of weakness: Protein structure poised at the edge of equilibrium. By using unnatural amino acids to manipulate the backbone hydrogen-bonding capacity, we reveal specific sites within LCDs that mediate self-association through the formation of labile cross-β structures. We found that multiple mutations causative of neurological disease target proline residues within the LCDs of functionally diverse proteins, and that these prolines serve to insulate labile cross-β structures and prevent aberrantly strong self-association. Denial of backbone-mediated hydrogen bonding at these mutation sites is sufficient to restore normal protein behavior.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: glen.liszczak@utsouthwestern.edu (G.L.); steven.mcknight@utsouthwestern.edu (S.L.M.)
Cite this article as X. Zhou *et al.*, *Science* **377**, eabn5582 (2022). DOI: 10.1126/science.abn5582

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abn5582

RESEARCH ARTICLE

BIOCHEMISTRY

Mutations linked to neurological disease enhance self-association of low-complexity protein sequences

Xiaoming Zhou[1], Lily Sumrow[1], Kyuto Tashiro[1], Lillian Sutherland[1], Daifei Liu[1], Tian Qin[1], Masato Kato[1,2], Glen Liszczak[1]*, Steven L. McKnight[1]*

Protein domains of low sequence complexity do not fold into stable, three-dimensional structures. Nevertheless, proteins with these sequences assist in many aspects of cell organization, including assembly of nuclear and cytoplasmic structures not surrounded by membranes. The dynamic nature of these cellular assemblies is caused by the ability of low-complexity domains (LCDs) to transiently self-associate through labile, cross-β structures. Mechanistic studies useful for the study of LCD self-association have evolved over the past decade in the form of simple assays of phase separation. Here, we have used such assays to demonstrate that the interactions responsible for LCD self-association can be dictated by labile protein structures poised close to equilibrium between the folded and unfolded states. Furthermore, missense mutations causing Charcot-Marie-Tooth disease, frontotemporal dementia, and Alzheimer's disease manifest their pathophysiology in vitro and in cultured cell systems by enhancing the stability of otherwise labile molecular structures formed upon LCD self-association.

Most genes of eukaryotic cells encode proteins that fold into stable, morphologically unique structures. Methods of protein structure determination deployed over the past six decades, including recent contributions from computer-enabled structure prediction, have resolved the shapes of a large fraction of the proteins that enable biological life. Knowledge of the molecular structures of proteins has been of considerable value in understanding many aspects of biology.

Between 10 and 20% of the proteome found in eukaryotic cells is unusual in failing to be capable of adopting a stable, three-dimensional structure (*1*). These proteins were first discovered as the activation domains of transcription factors (*2*, *3*), and differ from normal proteins by virtue of using no more than a few of the 20 amino acid residues found in normal proteins. These unusual protein domains have been described as being of low sequence complexity. The activation domain of the SP1 transcription factor, as an example, is composed almost exclusively of glutamine residues (*4*). Knowing that normal proteins rely on the chemical complexity of many different types of amino acid residues to adopt their specific, three-dimensional shape, it is not surprising that transcriptional activation domains fail to adopt stable molecular structures.

Since their discovery three decades ago in studies of gene-specific transcription factors, it has been shown that low-complexity domains (LCDs) are widespread. They decorate both ends of all 75 intermediate filament proteins found in mammals, fill the central channel of nuclear pores, adorn almost all RNA-binding proteins, and occur on the cytoplasmic faces of integral membrane proteins associated with mitochondria, neuronal vesicles, peroxisomes, lysosomes, and the Golgi apparatus. Whereas LCDs constitute no more than 10 to 20% of the proteomes of eukaryotic organisms, they serve as the reservoir for 75% of the combined forms of posttranslational modification used by cells to regulate protein function (*5*, *6*). Unfolded LCDs allow optimal access to the enzymes responsible for both depositing and eliminating marks of posttranslational modification.

Over the past two decades, numerous reports have revealed the capacity of LCDs to self-associate in a manner leading to their phase separation from aqueous solution (*7*–*11*). Assays of phase separation have facilitated mechanistic studies probing how LC domains interact. Here, we report two complementary approaches to studying how LCDs self-associate. First, we describe biochemical experiments on the LCD of an RNA-binding protein designated TAR DNA-binding protein 43 (TDP-43). Second, we investigate 10 idiosyncratic mutations causative of human neurological disease that map within LCDs.

TDP-43 is an RNA-binding protein involved in the movement of mRNA from the nucleus to cytoplasmic sites of translation. A growing list of mRNAs have been found to be translated in a localized manner, such that newly formed proteins are produced in locations proximal to their intended sites of need (*12*). During the process of movement from nuclear synthesis to cytoplasmic translation, mRNAs can be transiently organized in granular particles visible by light microscopy (*12*). These RNA granules are not encased within lipid membranes, and there is considerable interest in understanding how they are organized and regulated.

TDP-43 has long been recognized as a component of neuronal RNA granules that move mRNA molecules along the dendrites of neurons to facilitate enhanced translation at active synapses (*13*–*15*). A role in pathology is indicated by the fact that TDP-43 is frequently found in an aggregated state in the brain tissue of patients suffering from neurodegenerative disease (*16*, *17*).

A C-terminal region of the TDP-43 protein has been implicated in both its partitioning into RNA granules and its aberrant aggregation associated with disease. The C-terminal 153 residues of TDP-43 represent an LC region thought to be intrinsically disordered (*18*). That the TDP-43 LCD may help facilitate RNA granule formation is evidenced by its ability to become phase separated from aqueous solution. Upon incubation at high concentration under conditions of neutral pH and physiologically relevant monovalent salts, the TDP-43 LCD quickly forms liquid-like droplets that mature into a gel-like state (*19*, *20*). This behavior is reminiscent of the LCDs associated with many other RNA-binding proteins that have likewise been implicated in the organization of RNA granules (*9*, *10*, *21*). The relationship of the TDP-43 LCD to pathologic aggregation has been supported by the discovery of mutant amino acid substitutions within it that are causative of neurodegenerative disease (*22*, *23*).

The LCD of TDP-43 is different in one conspicuous regard from prototypical LCDs that phase separate. The latter proteins are routinely enriched with tyrosine or phenylalanine residues functionally important for phase separation (*7*, *8*, *24*–*26*). However, instead of being enriched in aromatic amino acids, the LCD of TDP-43 contains 10 evolutionarily conserved methionine residues. This feature imparts oxidation sensitivity to phase separated liquid-like droplets and hydrogels formed from the TDP-43 LCD (*19*). When exposed to H_2O_2, TDP-43 droplets melt in a manner correlated to the formation of methionine sulfoxide adducts. Upon chemical reduction by methionine sulfoxide reductase enzymes, thioredoxin, thioredoxin reductase, and NADPH, the TDP-43 LCD again coalesces into phase-separated droplets. This form of posttranslational modification of the TDP-43 LCD may define a redox switch allowing locally controlled dissolution of RNA granules (*19*).

A segment of 25 amino acids defines the region of the TDP-43 LCD critical for self-association and phase separation. This region

[1]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.
[2]Institute for Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology, Inage, Chiba, 263-8555, Japan.
*Corresponding author. Email: glen.liszczak@utsouthwestern.edu (G.L.); steven.mcknight@utsouthwestern.edu (S.L.M.)

has been perfectly conserved through the 400 million years of evolutionary divergence between fish and humans (19). It has also been observed by chemical footprinting to assume a structurally ordered state in phase-separated liquid-like droplets, hydrogels, and living cells (19). The same region of the TDP-43 LCD has been reported to form the molecular core of a morphologically specific cross-β structure (27).

The molecular assembly enabling self-association of the TDP-43 LCD, like all cross-β structures, should be reliant on intermolecular hydrogen bonds to help anneal adjacent β strands. If the role of main-chain hydrogen bonding is essential to self-association and phase separation of the TDP-43 LCD, then it can be concluded that self-association and phase separation result from the formation of labile structural order. An understanding of structural order should in turn allow for an understanding of the effects of subtle, disease-causing mutations within LCDs. Here, we report the results of experiments designed to test these assumptions.

Results

A purified protein fragment corresponding to the intact LC domain of TDP-43 forms phase-separated liquid-like droplets within seconds after suspension at room temperature in a neutral pH buffer supplemented with 150 mM NaCl. Compared with denatured, unfolded protein, residues M322 and M323 are protected from H_2O_2-mediated oxidation in phase-separated preparations of the TDP-43 LCD (19). These oxidation-resistant methionine residues colocalize with a dagger-like cross-β structure composed of 16 residues observed in three cryo–electron microscopy (cryo-EM) reconstructions of polymers formed from the TDP-43 LCD (27). The dagger-like structure and locus of oxidation protection themselves colocalize with a stretch of 25 residues corresponding to the most evolutionarily conserved region of the TDP-43 LCD (19).

Similar patterns of oxidation protection of residues M322 and M323 of the TDP-43 LCD have been observed in liquid-like droplets, labile cross-β polymers, and living cells (19). On the basis of the colocalization of these two methionine residues within the three structures resolved by cryo-EM, we hypothesized that self-association and consequential phase separation of the TDP-43 LCD might be nucleated by an assembly defined by the small cross-β structure.

Individual, main-chain hydrogen bonds are important for TDP-43 self-association

To experimentally investigate this hypothesis, we prepared 23 variants of the TDP-43 LCD differing only in chemical modification of the peptide backbone nitrogen atom of each residue within the ultraconserved region of the protein (Fig. 1A). Variants carried an N^{α}-methyl group on individual peptide backbone nitrogen atoms, thus enabling a systematic analysis of the involvement of main-chain hydrogen bonding in self-association and phase separation of the protein into liquid-like droplets. Cryo-EM structures of TDP-43 LCD polymers show that interpeptide hydrogen bonds represent favorable adhesive forces holding protein chains together in a parallel, in-register cross-β conformation. Our experimental strategy allowed assessment of whether hydrogen bonds located either within or outside of the cross-β structure might be required for phase separation.

Variants of the TDP-43 LCD methyl capped at individual backbone nitrogen atoms were assembled using a three-piece native chemical ligation strategy. Briefly, synthetic peptide thioesters bearing methyl-capped backbone nitrogen atoms at individual sites were inserted between flanking fragments of the TDP-43 LCD that had been expressed in bacterial cells (Fig. 1B). Ligation products were purified by high-performance liquid chromatography (HPLC) and analyzed by mass spectrometry to confirm the addition of a single backbone methylation site (Fig. 1C). The only amino acid not evaluated in this manner was residue P320. Proline, which has a side chain that is twice connected to the peptide backbone to form a pyrrolidine ring, is the only amino acid devoid of the NH group required for main-chain hydrogen bonding.

The native TDP-43 LCD sample and all 23 variants were purified and assayed for the formation of phase-separated liquid-like droplets in the presence of standard buffer, as well as buffers supplemented with 0.4, 0.8, or 1.2 M urea. Droplet formation was quantified in triplicate using a spectroscopic assay of turbidity and microscopic images of all proteins as observed in normal buffer, and the three different concentrations of urea are shown in fig. S1. As shown in Fig. 1D and E, the intact LCD of TDP-43 formed liquid-like droplets in native buffer. Fewer droplets were observed with 0.4 M urea, fewer still with 0.8 M urea, and no liquid-like droplets were observed in buffer supplemented with 1.2 M urea.

Compared with the native LCD of TDP-43, the variant bearing a methyl cap on the peptide backbone nitrogen associated with residue A324 formed fewer droplets in native buffer, fewer droplets in buffer supplemented with 0.4 M urea, and no droplets in buffer supplemented with either 0.8 or 1.2 M urea. Thus, the variant bearing a methyl cap on the peptide backbone nitrogen associated with alanine 324 was diminished in its capacity to phase separate under the three conditions supportive of droplet formation by the intact TDP-43 LCD. By contrast, a variant of the TDP-43 LCD bearing a methyl cap on the peptide backbone nitrogen associated with residue M337 formed morphologically indistinguishable droplets compared with the native protein under all conditions tested. Droplet numbers quantified by turbidity for the 337meM variant were indistinguishable from the native protein.

In this way, spectroscopic measurements for phase separation and microscopic images were gathered for all 23 N^{α}-methylated TDP-43 LCD variants. The turbidity assays displayed in Fig. 2 revealed nine variants as being impaired in their ability to phase separate, two variants as being partially impaired, and 12 variants as phase separating in a manner indistinguishable from the native protein. The histograms in Fig. 2C show normalized turbidity measurements for the 24 protein samples (y axis) relative to residue positions along the protein chain (x axis). One histogram shows normalized turbidity data collected in the absence of urea, one shows data collected in the presence of 0.4 M urea, and one shows data collected in the presence of 0.8 M urea. All three histograms reveal that all nine of the significantly impaired variants cluster between residues A321 and A329. The data shown in Fig. 2, D and E, further reveal that the region in which removal of single hydrogen bonds impedes phase separation colocalizes with both the footprinted region of structure-dependent oxidation protection, as well as the cross-β structure resolved by cryo-EM (19, 27).

A molecular structure has been determined for a hexapeptide segment of the TDP-43 LCD spanning residues 321 to 326 (28). This hexapeptide forms parallel β-sheets wherein strands are held together by main-chain hydrogen bonds. Each of these six residues, if bearing a methyl group on their peptide backbone nitrogen in the context of the intact TDP-43 LCD, exhibited a weakened propensity to phase separate into liquid-like droplets (Fig. 2).

Having analyzed the 23 N^{α}-methylated TDP-43 LCD variants, we prepared samples in which three peptide backbone nitrogen atoms were concomitantly methylated. One such variant capped the peptide backbone nitrogen atoms of residues A321, M323, and A325; the other capped the peptide nitrogen atoms of residues M323, A325, and Q327. The same methods of peptide synthesis and three-piece chemical ligation described in Fig. 1 were used to assemble these variants of the TDP-43 LCD. As shown in Fig. 3, both variants were severely impeded with respect to the formation of phase-separated liquid-like droplets. No droplets were observed for either variant in the presence of any concentration of added urea, and the number of droplets formed in the absence of urea was significantly reduced.

By devising and deploying methods of semisynthetic protein chemistry, we have systematically evaluated the importance of 23 peptide backbone amide protons within the TDP-43 LCD for the formation of phase-separated

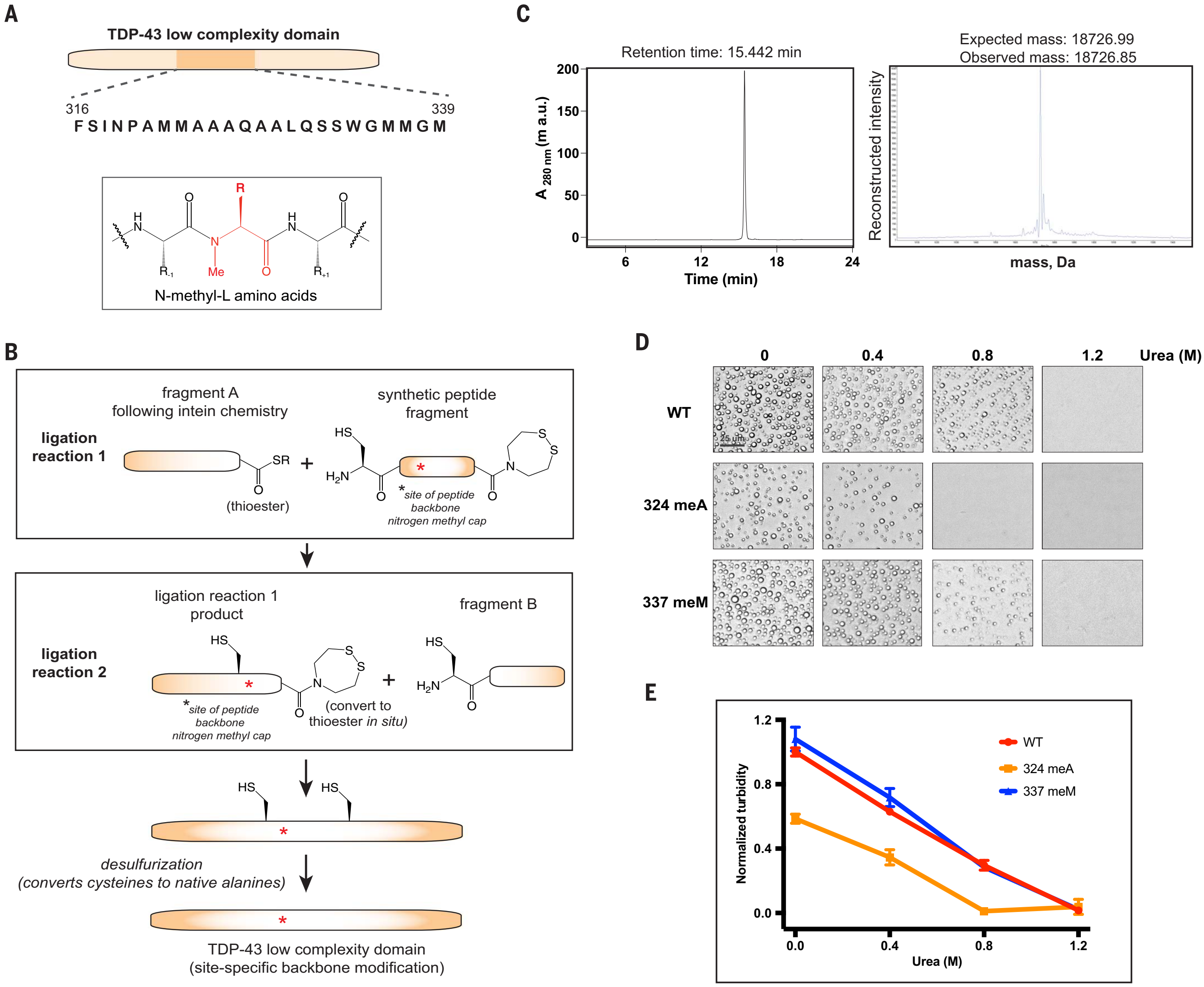


Fig. 1. Preparation of semisynthetic derivatives of the TDP-43 LCD carrying single, methyl-capped peptide backbone nitrogen atoms. (**A**) Schematic representation of the LCD of TDP-43 (residues 262 to 414) with an ultraconserved region (residues 316 to 339) highlighted by darker shading. N^{α}-methyl amino acid (red) scanning analysis was applied on the ultraconserved region. (**B**) Schematic depicting the strategy to prepare N^{α}-methyl amino acid (red asterisk)–incorporated TDP-43 LCDs. N^{α}-methyl amino acids were introduced into synthetic peptides that were conjugated by sequential native chemical ligation reactions to flanking protein fragments produced in bacteria. Ligation products contained cysteine residues in place of alanine residues at both ligation boundaries. Both cysteine residues were chemically desulfurized to alanine such that the only difference between engineered LCDs and the native LCD was the presence of a single N^{α}-methyl group at the desired site. (**C**) Representative result of HPLC trace (left) and high-resolution, intact MS (right) for confirming ligation product from (B). (**D**) Light microscopic analyses of phase separation for three representative samples, including the reconstructed, semisynthetic, native TDP-43 LCD (WT), a variant carrying a methyl cap on the peptide backbone nitrogen atom of alanine residue 324 (324 meA), and a variant carrying a methyl cap on the peptide backbone nitrogen of methionine residue 337 (337 meM). Each protein was incubated under conditions of neutral pH and physiologically normal monovalent salts, allowing for the formation of phase-separated liquid-like droplets. Droplet formation was assayed in triplicate in normal buffer and in buffers supplemented by 0.4, 0.8, and 1.2 M urea. Scale bar, 25 μm. (**E**) Quantitative analysis of phase separation for all protein samples monitored by spectroscopic analysis of turbidity.

liquid-like droplets. Denial of hydrogen-bonding capacity in a localized region of the TDP-43 LCD distinctly weakened protein self-association. This region of sensitivity to the focused removal of hydrogen bonds falls within a known cross-β structure. Therefore, our observations comport with the proposed involvement of the small cross-β structure in self-association and phase separation by the TDP-43 LCD. These observations further conform to previous studies of the fused-in-sarcoma (FUS) and heterogeneous nuclear RNPA2 (hnRNPA2) LCDs, in which labile cross-β structures also play a role in phase separation (*9*, *29*).

The experimental observations reported in this study give evidence of the importance of main-chain hydrogen bonding for self-association and phase separation of the TDP-43 LCD. That elimination of single hydrogen bonds can reproducibly yield distinct phenotypic deficits gives evidence of the inherent

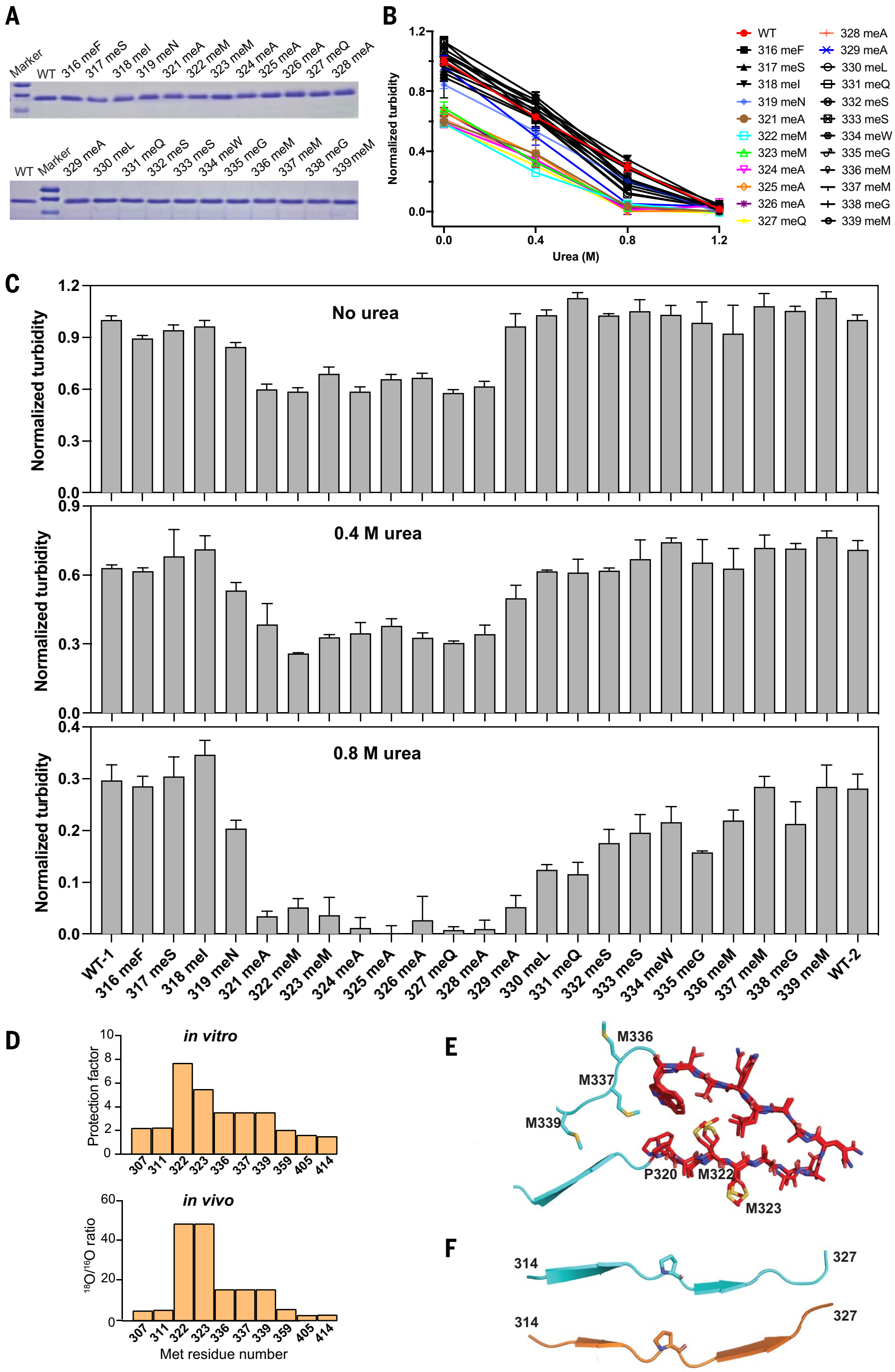


Fig. 2. Phase-separation capacities of 23 variants of the TDP-43 LCD differing according to the presence of a single, methyl-capped peptide backbone nitrogen atom. (**A**) Purity and integrity evaluation of all N$^{\alpha}$-methyl variants of the TDP-43 LCD by SDS-PAGE. (**B**) Normalized turbidity measurement of TDP-43 LCD phase separation induced in buffer of neutral pH and physiological monovalent salt, or in buffers supplemented by 0.4, 0.8, and 1.2 M urea. All samples were analyzed in triplicate and further evaluated and photographed by light microscopy (fig. S1). Turbidity measurements revealed two classes of variants irrespective of the presence or absence of urea. One class composed of nine variants showed, relative to the native TDP-43 LCD, reduced turbidity and fewer liquid-like droplets under all conditions. The other class, composed of 13 variants, showed evidence of phase separation indistinguishable from the native protein. (**C**) Histograms reveal the normalized, average turbidity measurements of 24 protein samples as evaluated in the absence of urea as well as in the presence of either 0.4 or 0.8 M urea. Turbidity is graphed on the *y*-axis of each plot relative to the amino acid sequence of the ultraconserved region of the TDP-43 LCD between residue F316 and residue M339 (*x*-axis). Irrespective of assay condition, all nine of variants exhibiting impediments to phase separation clustered between residues A321 and A329. (**D**) Reproductions of published data pinpointing the region of the TDP-43 LCD footprinted to reveal structure-dependent protection from methionine oxidation (*19*). (**E**) Molecular structure of cross-β polymers formed from the TDP-43 LCD as resolved by cryo-EM (*27*). The region of the protein that was observed to be of the same molecular structure in three independent cryo-EM reconstructions extends from residue P320 to residue M337. (**F**) Top image shows ribbon diagram representations of residues 314 to 327 from one of the three cryo-EM structures of cross-β polymers made from the TDP-43 LCD shown in (E) (*27*). Bottom image shows ribbon diagram representation of residues 314 to 327 from a different cryo-EM structure of cross-β polymers made from the intact TDP-43 LCD (*32*).

lability of the cross-β structure responsible for self-association of the TDP-43 LCD.

Several studies have suggested that the TDP-43 LCD can adopt an α-helical conformation in the same region that we have shown to function through the formation of labile cross-β interactions (*18*, *30*, *31*). Evidence of α-helical conformation has been observed by solution [nuclear magnetic resonance (NMR)] spectroscopy under conditions of acidic pH and in the absence of monovalent salt. Two such studies have shown that when pH is raised to neutral, NMR spectra diagnostic of α-helical conformation disappear and are replaced by spectra diagnostic of β-strand conformation (*18*, *30*). Thus the α-helical conformation is likely not stable under physiological conditions of pH and salt, as used in our experiments.

Amino acid side chains are important to self-association of the TDP-43 LCD

Whereas data presented in Figs. 1, 2, and 3 give evidence that main-chain hydrogen bonding is important for self-association of the TDP-43

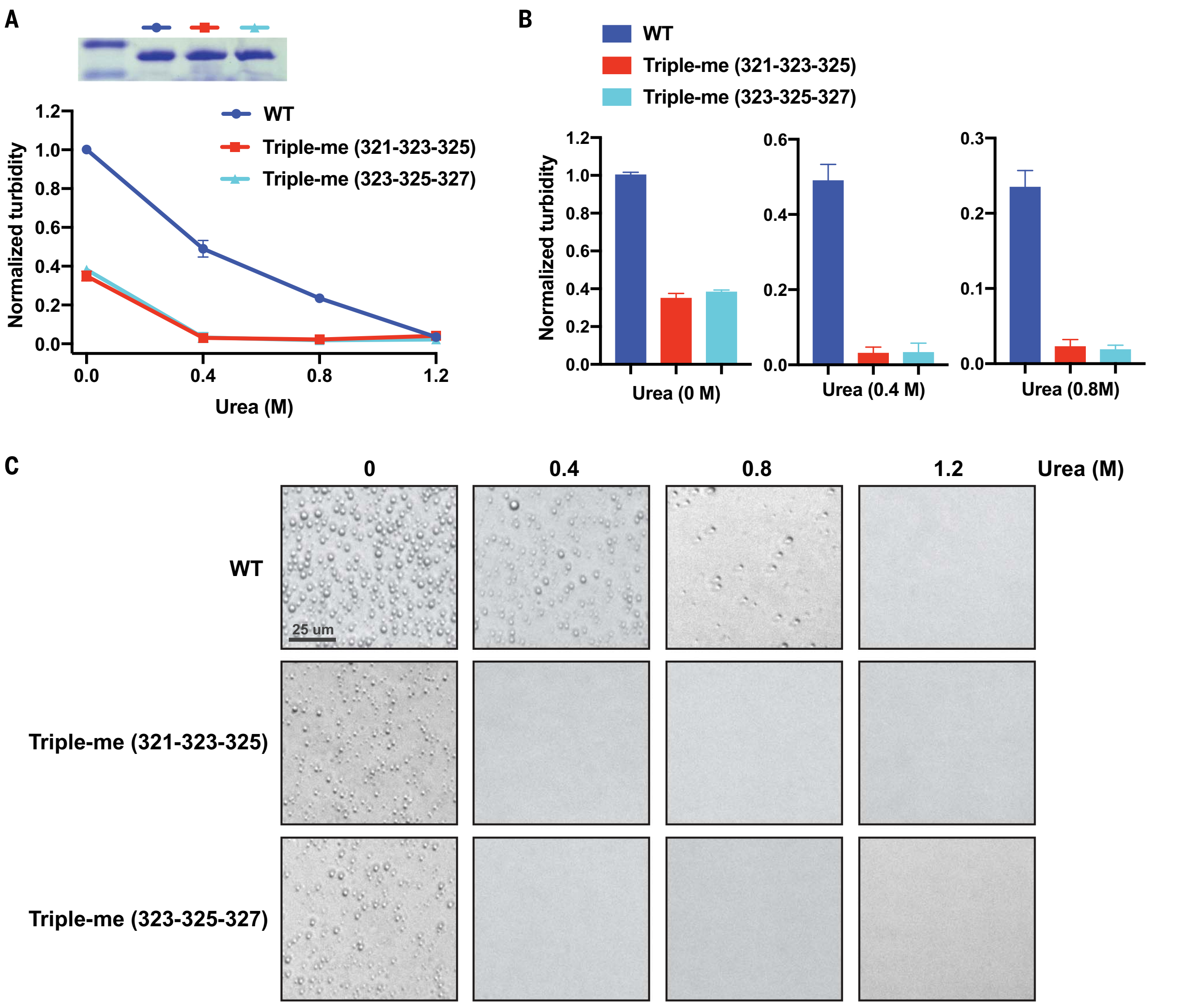


Fig. 3. Preparation and analysis of semisynthetic derivatives of the TDP-43 LCD carrying three methyl-capped peptide backbone nitrogen atoms. (**A**) Synthetic peptides carrying methyl caps on the peptide backbone nitrogen atoms of residues A321, M323, and A325 [triple-me (321-323-325)] or residues M323, A325, and Q327 [triple-me (323-325-327)] were inserted into the full-length LCD of TDP-43 by chemical ligation as described in Fig. 1. Purity and integrity of the ligation products were evaluated by SDS-PAGE (top panel). Phase separation of the ligation products was quantified by turbidity measurement (bottom panel). The phase separation was induced in buffer of neutral pH and physiological monovalent salt supplemented with no urea or 0.4, 0.8, or 1.2 M urea. (**B**) Presentation of normalized turbidity measurements in the form of histograms. No evidence of turbidity above background was observed for either sample bearing three methyl-capped peptide backbone nitrogen atoms when assayed in the presence of 0.4, 0.8, or 1.2 M urea. In the absence of supplemented urea, both samples bearing three methyl-capped nitrogen atoms displayed turbidity levels reduced by between 60 and 70% relative to that of the native TDP-43 LCD. (**C**) Light microscopic analyses of phase separation for native TDP-43 LCD (WT), triple-me (321-323-325), and triple-me (323-325-327) variants assayed in (A) and (B). Scale bar, 25 μm.

LCD, they fail to address the potential importance of amino acid side chains. In the absence of side chain–mediated chemical interactions, there would be no opportunity for specificity in formation of the cross-β structure itself. To investigate the functional involvement of amino acid side chains, we performed glycine-scanning mutagenesis across the ultraconserved region of the TDP-43 LCD. We chose glycine because of the chemical simplicity of its side chain.

The pattern of effects of 26 side chain variants from amino acids 315 to 340 of the TDP-43 LCD upon phase separation was similar to the pattern of effects resulting from hydrogen bond removal (Fig. 4). Substitution of any native residue between A324 and L330 by glycine led to the formation of either amorphous precipitates or distinctly misshapen droplets. This region of the protein was also most sensitive to methyl capping of individual peptide backbone nitrogen atoms (Fig. 2).

Of the 10 variants located on the C-terminal side of L330, only two caused notable changes in droplet morphology (G335S and G338S). These are the only positions within the region chosen for study that are represented by glycine in the native sequence of TDP-43 (and were changed to serine as part of the mutagenesis scan). Of the nine variants located on the N-terminal side of A324, only two caused notable changes in phase separation. Both the N319G and P320G variants led to the formation of

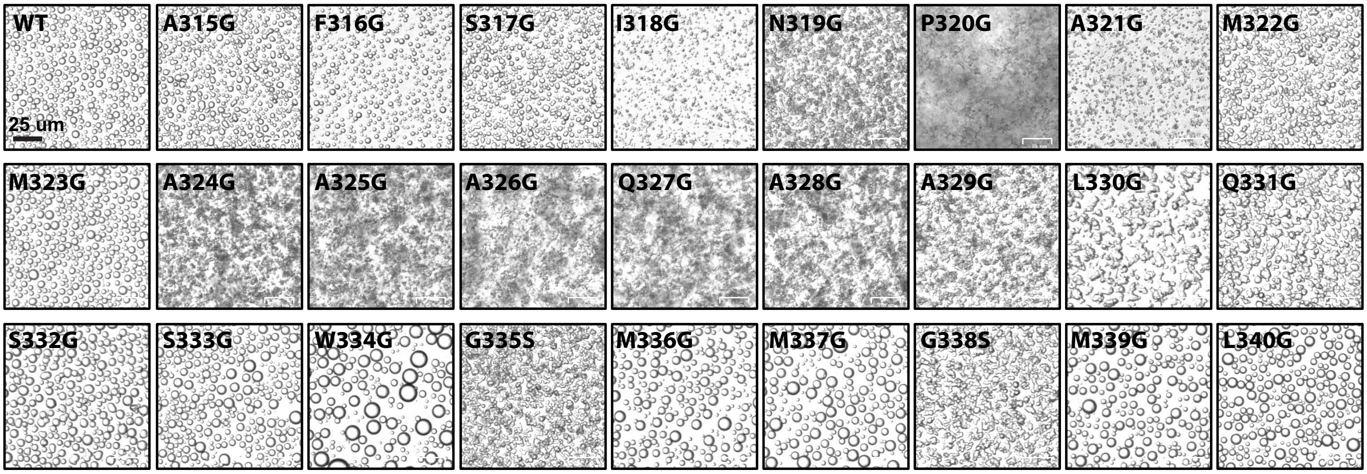


Fig. 4. Phase-separation capacities of 26 variants of the TDP-43 LCD differing according to the replacement of a single amino acid residue with glycine. Single amino acid residues within the ultraconserved region of the TDP-43 LCD were changed to glycine by conventional mutagenesis and expression in bacterial cells. Two of the 26 positions already contained glycine (G335 and G338) in the native sequence and were mutationally changed to serine. Each protein was purified and tested for phase separation in a buffer of neutral pH and physiologically normal monovalent salt ions. Light microscopic images were photographed 1 hour after the initiation of each reaction. In addition to assays performed in standard buffer, each protein sample was also evaluated for its capacity to phase separate in the presence of 0.4, 0.8, and 1.2 M urea (fig. S2). Scale bar, 25 μm.

amorphous precipitates. Among all 26 variants tested, the P320G variant revealed the most profound effect on droplet formation. Immediately after suspension in aqueous buffer, the P320G variant formed a heavily tangled precipitate.

In addition to inspecting the 26 glycine-scanning variants in normal aqueous buffer, we analyzed droplet morphology in buffers supplemented with 0.4, 0.8, and 1.2 M urea. As shown in fig. S2, the pattern of effects of glycine-scanning variants was largely similar irrespective of the addition of urea. These added assays revealed that certain variants were, relative to the native protein, either more or less sensitive to the urea denaturant. In particular, the four "outlying" variants, N319G, P320G, G335S, and G338S, continued to display either precipitates or misshapen droplets in the presence of 1.2 M urea (a condition that fully melts liquid-like droplets made from the native TDP-43 LCD). In summary, these experiments confirm the importance of several amino acid side chains within the ultraconserved region of the TDP-43 LCD for the formation of labile and morphologically spherical, phase-separated droplets.

Residue P320 is important for the proper balance of TDP-43 self-association

Having observed profound deficits in phase-separation assays for the P320G variant, we wondered whether mutation of other proline residues within the TDP-43 LCD might also lead to significant deficits. In addition to P320, the sequence of the LCD specifies residues at positions P280, P349, and P363 outside of the ultraconserved region (Fig. 5A). Each of these residues was individually changed to glycine, followed by expression, purification, and assay of phase separation. None of the three variants caused any effect on either the morphology or the stability of phase-separated liquid-like droplets (Fig. 5B). We thus conclude that residue P320 is of particular importance to the function of the TDP-43 LCD.

A number of well-validated, amyotrophic lateral sclerosis (ALS)–causing missense mutations map within or close to the ultraconserved region of the TDP-43 LCD (*23*). Included among such mutations is the particularly well-studied M337V variant that leads to the formation of oxidation-resistant liquid-like droplets (*19*). These observations raise the question of why human genetic studies have failed to report ALS-causing missense mutations in residue P320?

Given its severe propensity to aggregate into stable precipitates, we considered the possibility that expression of the P320G variant might be incompatible with cell viability. To test this, U2OS cells were stably transformed with inducible transgenes encoding native TDP-43, the M337V ALS-causing variant, or the P320G variant. After isolating cell clones expressing comparable levels of the three proteins upon conditional induction, and seeing no detectable expression of the proteins in the absence of the chemical inducer, we performed viability studies as a function of time after induction. As shown in Fig. 5C, conditionally induced expression of the P320G variant substantially reduced cell viability. No such impediment to viability was observed for cells transformed with the parental expression vector, nor for cells conditionally expressing either the native TDP-43 protein or the ALS-causing M337V variant. We conclude that the P320G variant has not appeared in human genetic studies as an ALS-causing missense mutation because it is cell lethal.

To investigate the behavior of the P320G variant in living cells, compared with the native protein, the ALS-disposing M337V variant, and the three P-to-G variants at positions 280, 349, and 363, expression vectors were prepared linking each variant of TDP-43 to green fluorescent protein (GFP). U2OS cells were transiently transfected with each expression vector. Twenty-four hours later, the cells were visualized by confocal microscopy as shown in Fig. 5D. GFP signal associated with the fusion protein linked to the native TDP-43 protein was restricted to the nuclei of transfected cells, and the P280G, P349G, and P363G variants yielded patterns of nuclear staining indistinguishable from the native TDP-43 protein. Similarly, we observed nuclear staining for the ALS-causing M337V variant. By contrast, the GFP signal for the P320G variant was almost exclusively restricted to the cytoplasm, with residual nuclear staining confined to prominent nuclear puncta.

Why does P320G show such an unusual phenotype? Recognizing that proline is the only amino acid not permissive of forming main-chain hydrogen bonds, its replacement by any other amino acid would obviously restore hydrogen-bonding capacity. In addition to the P320G variant characterized thus far, we also mutated proline 320 to either serine or alanine. Both variants caused equally profound precipitation of the TDP-43 LCD relative to the P320G variant in phase-separation assays (fig. S3). These observations suggest that

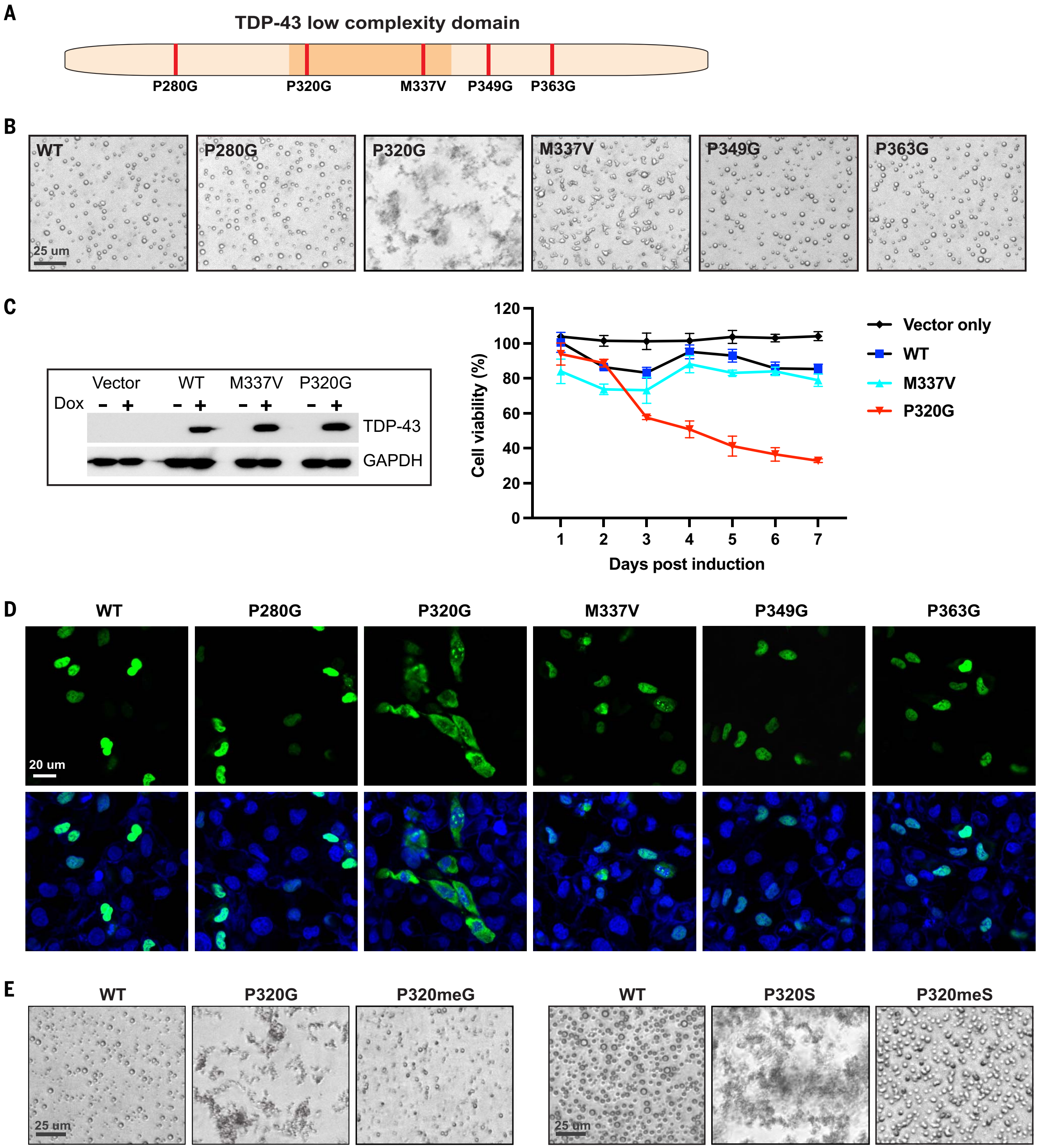


Fig. 5. Evidence of the particular importance of residue P320 of the TDP-43 LCD. (**A**) Schematic depicting the positions of four proline residues (P280, P320, P349, and P363) and the ALS-causing M337V variant within the LCD of TDP-43. Darker shading region corresponds to the ultraconserved region. (**B**) Protein samples corresponding to the native LCD of TDP-43 (WT), four proline-to-glycine variants (P280G, P320G, P349G, and P363G), and the M337V variant were assayed for the formation of liquid-like droplets. Other than the P320G variant, which formed aggregated tangles, all other proline-to-glycine variants generated spherical droplets indistinguishable from those made from the native TDP-43 LCD. The M337V variant maintained phase-separation capacity but yielded misshaped droplets. Scale bar, 25 μm. (**C**) U2OS cells were stably transformed with vectors allowing for conditional, doxycycline-mediated expression of FLAG-tagged versions of native, full-length TDP-43 (WT), the ALS-causing M337V variant of TDP-43, or the P320G variant causative of severe precipitation. Single transformant clones were isolated in the absence of doxycycline and screened for the purpose of finding stable clones that expressed equivalent protein levels after doxycycline induction. Cell lysates were recovered before and 24 hours after doxycycline induction and Western blotted using a FLAG antibody (left panel). Cell viability was monitored daily after induction for each clone, revealing minimal effects from doxycycline-induced expression of

either the native TDP-43 protein (WT) or the ALS-causing M33V variant compared with substantial deficits resulting from expression of the P320G variant (right panel). (**D**) Constitutive expression vectors encoding fusion proteins linking GFP to the N terminus of full-length TDP-43 were transiently transfected into U2OS cells. Nucleus-restricted GFP was observed for native TDP-43 linked to GFP, as well as for the proline-to-glycine variants of residues 280, 349, and 363 and the ALS-causing M337V variant. Transient expression of the P320G variant led to predominantly cytoplasmic GFP staining, with residual nuclear staining appearing aberrantly punctate. Scale bar, 20 μm. (**E**) Semisynthesis of variants of the TDP-43 LCD, P320G and P320S, formed tangled precipitates upon tests of phase separation. Methyl capping of the peptide backbone nitrogen atom associated with the glycine residue of the P320G variant, or the peptide backbone nitrogen atom associated with the serine residue of the P320S variant, yielded proteins for which formation of phase-separated liquid-like droplets was restored. Scale bar, 25 μm.

P320 might function as a chemical insulator functioning to locally buffer the extent of cross-β interactions. This is consistent with the location of P320 between two adjacent β strands, as revealed both by the dagger structure resolved by Eisenberg and colleagues and by an independent cryo-EM structure recently deduced for polymers formed from the intact TDP-43 LCD (Fig. 2F) (*27*, *32*). These independently resolved structures are virtually identical for the region of the TDP-43 LCD spanning residues 314 to 327.

If the glycine or serine substitutions of residue P320 increase TDP-43 aggregation propensity by regenerating main-chain hydrogen bonds, then this should be reversed by methylating the peptide backbone nitrogen atom of the glycine residue in P320G or that of the serine residue in P320S. We again used the TDP-43 synthesis strategy detailed in Fig. 1 to introduce a methyl cap onto the peptide backbone nitrogen atom of these two variant residues. In both cases, these efforts led to the rescue of phase-separated liquid-like droplets (Fig. 5E).

Mutational variants of evolutionarily conserved proline residues in NFL, tau, and hnRNPA2 enhance the stability of otherwise labile cross-β interactions

The effects of the glycine, serine, and alanine missense variants of proline 320 within the TDP-43 LCD prompted us to search for disease-causing mutations within LCDs that might map to proline residues. We found idiosyncratic disease-associated proline mutations within LCDs of the neurofilament light (NFL) chain protein, the microtubule-associated protein tau, and the hnRNPA2 protein. Each of these proteins is known to be subject to autosomal-dominant mutations of proline residues causative of disease, and each became the focus of our biochemical approach to the study of LCDs.

Mutations changing residue P8 or P22 of NFL cause CMT disease

Human Charcot-Marie-Tooth (CMT) disease can be caused by mutation of either of two proline residues located within the head domain of the NFL protein (*33*, *34*). These autosomal-dominant mutations change residue P8 to leucine, arginine, or glutamine or residue P22 to serine, threonine, or arginine. Various assays have been used to confirm that these mutations impede the proper assembly of neurofilaments in living cells (*33*, *35*, *36*).

Intermediate filament (IF) proteins can be expressed in bacterial cells, purified, and incubated under conditions permissive of formation of mature filaments. The first two steps in assembly, dimerization and tetramerization, are mediated solely by centrally located, α-helical rod domains, and can proceed in the absence of the head domain. By contrast, the head domains of IF proteins are essential for the subsequent assembly of eight tetramers into cylindrical filaments (*37*–*39*).

The head domains of IF proteins are of low sequence complexity and have long been thought to function in the absence of molecular structure (*40*, *41*). Alternative observations have now been reported for the head domains of the desmin, vimentin, peripherin, neurofilament light chain, neurofilament medium chain, and neurofilament heavy chain IF proteins. Isolated protein samples of each of these IF head domains have been found to phase separate through the formation of labile cross-β structures (*42*). Thus, IF head domains share functional relatedness with the LCDs of certain RNA-binding proteins that self-associate in the form of liquid-like droplets and hydrogels (*8*, *9*).

A solid-state NMR spectroscopy study showed that the head domains of the desmin and NFL intermediate filament proteins form labile cross-β structures (*43*). Intein ligation enabled segmental isotopic labeling, showing that the NMR spectra of the isolated desmin and NFL head domains match the spectra observed in fully assembled intermediate filaments. Such studies are inconsistent with the notion that IF head domains function in the absence of molecular structure, and instead showed that they form labile and reversible cross-β interactions. This concept of LCD self-association conforms to similar studies on the LCDs of hnRNPA2, yeast ataxin-2, FUS (*9*, *29*, *44*), and our findings here on TDP-43.

The recent Zhou *et al.* (2021) study of IF head domains also evaluated human mutations in desmin causative of cardiac deficits and in NFL causative of CMT disease (*43*). All desmin and NFL head domain mutations were found to enhance the stability of otherwise labile cross-β interactions. Such observations comport with the autosomal-dominant nature of the human mutations in desmin and NFL, and suggest that aberrantly enhanced head domain self-association is incompatible with proper intermediate filament assembly.

The proximity of the P8 and P22 CMT mutations to the observed position of the labile cross-β structure within the NFL head domain is reminiscent of findings for the P320G, P320S, and P320A variants of the TDP-43 LCD (Fig. 5 and fig. S3). We reasoned that, as for P320 in TDP-43, residues P8 and P22 within the NFL head domain might locally buffer against the aberrant spread of cross-β structure. Consequently, mutations at these positions, which insulate cross-β structures, result in pathologic strengthening of head domain self-association. If this is true, then aberrant self-association may be reversed by eliminating the added hydrogen bond in mutant constructs. This idea was tested through methyl capping of the peptide backbone nitrogen atoms associated with the leucine, arginine, and glutamine variants of P8, as well as the serine, threonine, and arginine variants of P22.

Methylation of the peptide backbone nitrogen of CMT-causing mutational variants restores normal NFL protein function

A native chemical ligation reaction was used to link synthetic peptides containing single, N^{α}-methyl amino acids to the remaining 517 residues of the NFL polypeptide (Fig. 6A). These semisynthetic NFL proteins were purified and incubated under conditions conducive to assembly of mature intermediate filaments. As shown in Fig. 6B, NFL proteins bearing any of the six CMT mutations formed amorphous tangles. By contrast, all six of the methyl cap–repaired variants formed morphologically normal intermediate filaments indistinguishable from those assembled from the native NFL protein.

The P8Q, P8L, P8R, P22R, P22S, and P22T N-terminal peptides were much more prone to precipitation than the cognate peptide bearing the native sequence. Unexpectedly, the methyl-capped versions of these six CMT-mutated variants were uniformly more soluble than the parental substrates bearing normal peptide nitrogen atoms at the P8Q, P8L, P8R, P22R, P22S, and P22T positions. We turned to acquisition of thioflavin-T (ThT) fluorescence

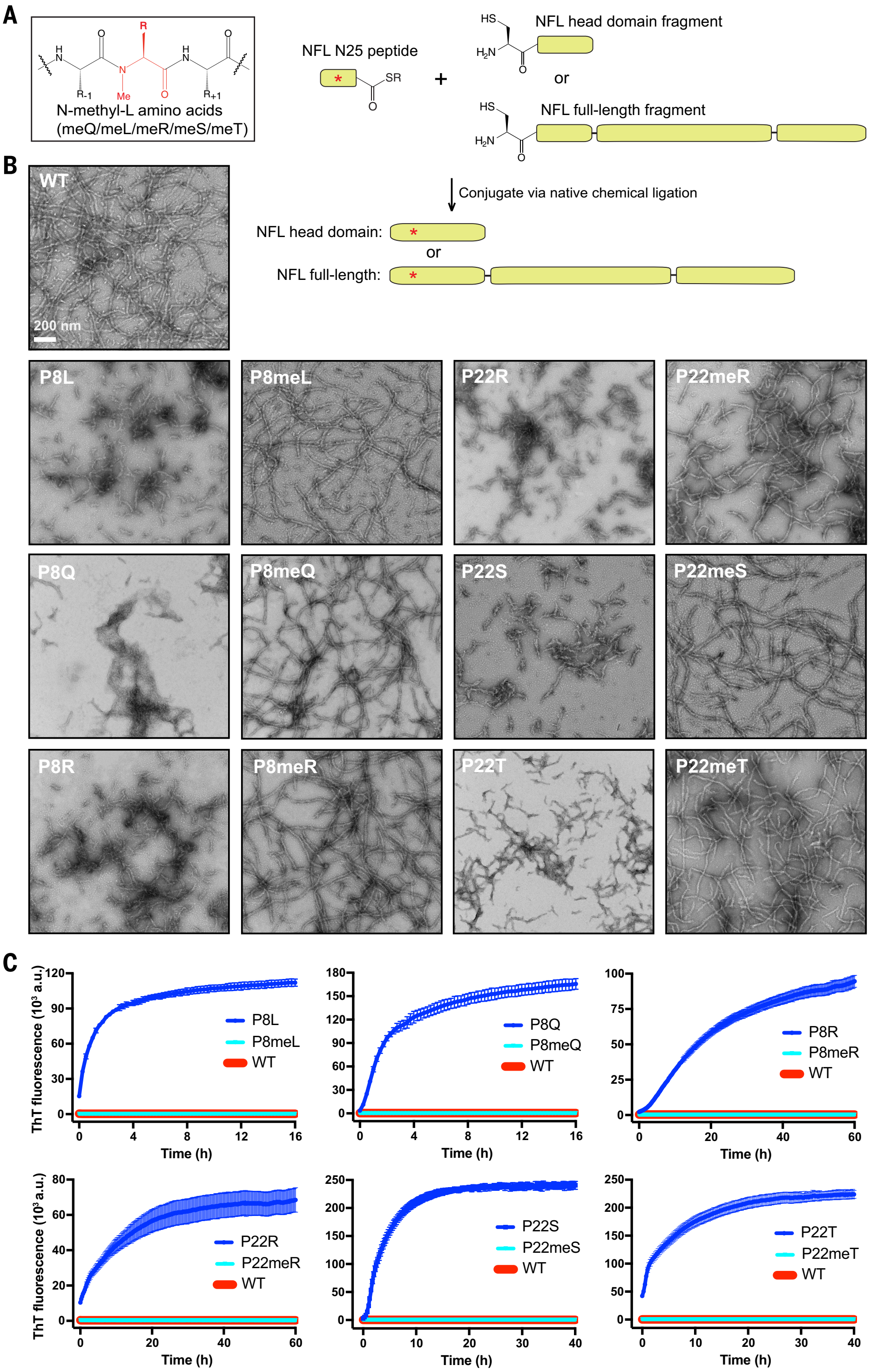


Fig. 6. Effects of methyl-capping peptide backbone nitrogen atoms associated with disease-causing residues within the NFL head domain. (**A**) Schematic representation of the strategy to prepare N$^{\alpha}$-methyl amino acid (red color, red asterisks)–incorporated NFL head domain and NFL full-length protein. Synthetic peptide thioesters corresponding to the N-terminal 25 residues of the NFL protein were prepared to contain (i) the native sequence of NFL, (ii) CMT mutational variants at position 8 or 22, or (iii) CMT variants methyl capped at the peptide backbone nitrogen in the mutated amino acid. Native chemical ligation was used to produce full-length NFL protein or the isolated head domain bearing different synthetic peptides at the N terminus. (**B**) Full-length NFL proteins carrying P8L, P8Q, and P8R variations; P22R, P22S, or P22T variations; or N$^{\alpha}$-methyl variants of these six residues were incubated under conditions receptive to the formation of intermediate filaments. Filament assembly was monitored by transmission electron microscopy. Assembly assays for all six CMT variants yielded tangled, amorphous precipitates. Assembly assays for semisynthetic CMT variants containing a methyl-capped nitrogen (P8meL, P8meQ, P8meR, P22meR, P22meS, and P22meT) yielded homogeneous intermediate filaments indistinguishable from those made from the native NFL protein (WT). Scale bar, 200 nm. (**C**) Synthetic peptides used to generate semisynthetic, full-length NFL proteins were assayed for polymerization as monitored by acquisition of ThT fluorescence. All six peptides bearing a CMT-causing lesion (P8L, P8Q, P8R, P22R, P22S, and P22T) yielded ThT curves indicative of time-dependent polymerization. No evidence of polymerization was observed for the parental peptide (WT) or for any of the six CMT-causing variants (P8meL, P8meQ, P8meR, P22meR, P22meS, and P22meT) that were modified to methyl cap the peptide backbone nitrogen atom of the variant amino acid residue.

to evaluate the kinetics of peptide aggregation (Fig. 6C). The peptide containing the native NFL sequence showed no evidence of polymerization, as revealed by a time-dependent increase in ThT fluorescence. By contrast, peptides containing each of the six CMT-causing variations within the NFL head domain showed prominent ThT curves. In the case of all six CMT mutants, methyl capping of the peptide backbone nitrogen atom associated with the variant amino acids completely eliminated peptide polymerization.

Three of the variant peptides, P8L, P8Q, and P8R, became insoluble immediately upon dilution into aqueous buffer, thus preventing assays of time-dependent polymerization. The problem of insolubility was overcome by performing ThT assays in the presence of either 1 M urea (for P8Q and P8R) or 2 M urea (for P8L). All other peptides were assayed in normal aqueous buffer.

Enhanced peptide polymerization (Fig. 6C) and aberrant IF assembly (Fig. 6B) were assumed to result from mutation-directed strengthening of NFL head domain self-association. To test this prediction, synthetic peptides carrying the P8Q and P22S mutations,

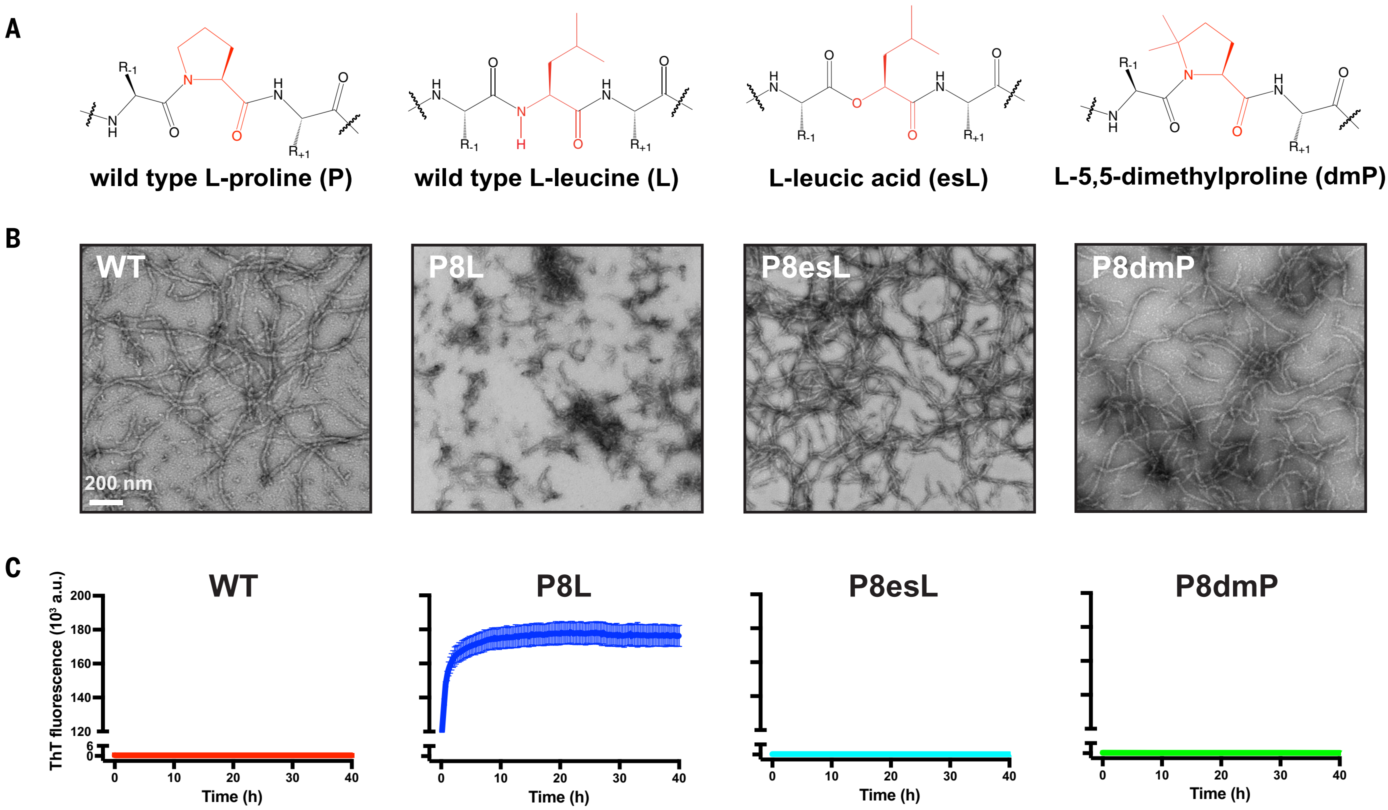


Fig. 7. Effects of eliminating main-chain hydrogen bonding using amide-to-ester substitutions in the peptide backbone and effects of replacing proline 8 of the NFL head domain with dmP. (**A**) Structures of synthetic peptides with chemical variation at residue P8 of the NFL head domain highlighted in red. Synthetic peptides were prepared by replacing residue P8 of the NFL head domain with leucine, L-leucic acid, or dmP. (**B**) Semisynthetic, full-length NFL proteins bearing the structures shown in (A) were assayed for assembly of bona fide intermediate filaments. Reconstruction of the P8L CMT variant yielded tangled, amorphous precipitates. Amide-to-ester backbone substitution at leucine residue 8 (P8esL), which thus eliminated a hydrogen bond donor from this site, facilitated assembly of intermediate filaments indistinguishable from those assembled from the native NFL protein. Replacement of residue P8 with dmP (P8dmP) allowed assembly of homogeneous intermediate filaments morphologically indistinguishable from those produced by the native protein (WT). Scale bar, 200 nm. (**C**) Synthetic peptides used to generate full-length NFL proteins were assayed for polymerization as monitored by acquisition of ThT fluorescence. The P8L CMT variant peptide exhibited clear evidence of polymerization not observed for the parental peptide bearing the native sequence of NFL (WT). The WT, P8esL, and P8dmP peptides showed no evidence of polymerization.

or N$^{\alpha}$-methyl derivatives thereof, were ligated to the remainder of the NFL head domain (Fig. 6A). Ligation products were purified and incubated under conditions known to distinguish between the solubility of the native head domain of NFL relative to CMT mutational variants (*43*). As shown in fig. S5, the native NFL head domain was soluble under such conditions, variants carrying either the P8Q or P22S mutations were not, and methyl capping of the peptide backbone nitrogen atoms associated with P8Q and P22S fully restored solubility.

Esterification of the peptide bond associated with the P8L CMT variant of the NFL head domain restores normal protein function

We propose that each of the three variants at the P8 position, as well as the three variants at the P22 position, cause the strengthening of an otherwise labile and transient cross-β structure useful for NFL head domain self-assembly. Enhanced strength of self-association is interpreted to result from an extended hydrogen bond network, and this extension is prevented by methyl capping of the peptide backbone nitrogen atoms associated with the variant amino acid residues. To further probe this concept, we used a different chemical approach to prevent hydrogen bonding by one of the variant amino acid residues. Instead of methyl capping the peptide backbone nitrogen associated with the leucine residue of the P8L CMT mutant, we replaced the peptide backbone nitrogen with oxygen (Fig. 7A).

A synthetic peptide bearing an amide-to-ester backbone substitution at the P8L site was ligated to the remainder of the NFL protein. The semisynthetic protein was purified and incubated under conditions suitable for the assembly of intermediate filaments. As shown in Fig. 7B, this amide-to-ester variant formed intermediate filaments indistinguishable from those assembled from the native NFL protein. Thus, esterification at the appropriate position of the polypeptide chain repaired the P8L deficit to an extent indistinguishable from methyl capping of the P8L peptide backbone nitrogen. We further observed that the amide-to-ester substitution fully solubilized the P8L synthetic peptide, as determined by time-dependent acquisition of ThT fluorescence (Fig. 7C).

dmP functionally replaces residue P8 of the NFL head domain

Other than being the sole amino acid incapable of forming main-chain hydrogen bonds, proline is also notable for its ability to adopt either the *cis* or *trans* peptide bond conformation. To investigate the potential importance of this chemical feature of proline, we replaced residue P8 of the NFL head domain with 5,5-dimethyl-L-proline (dmP) (Fig. 7A). Relative to the normal proline side chain, dmP strongly prefers the *cis* conformation (*45*).

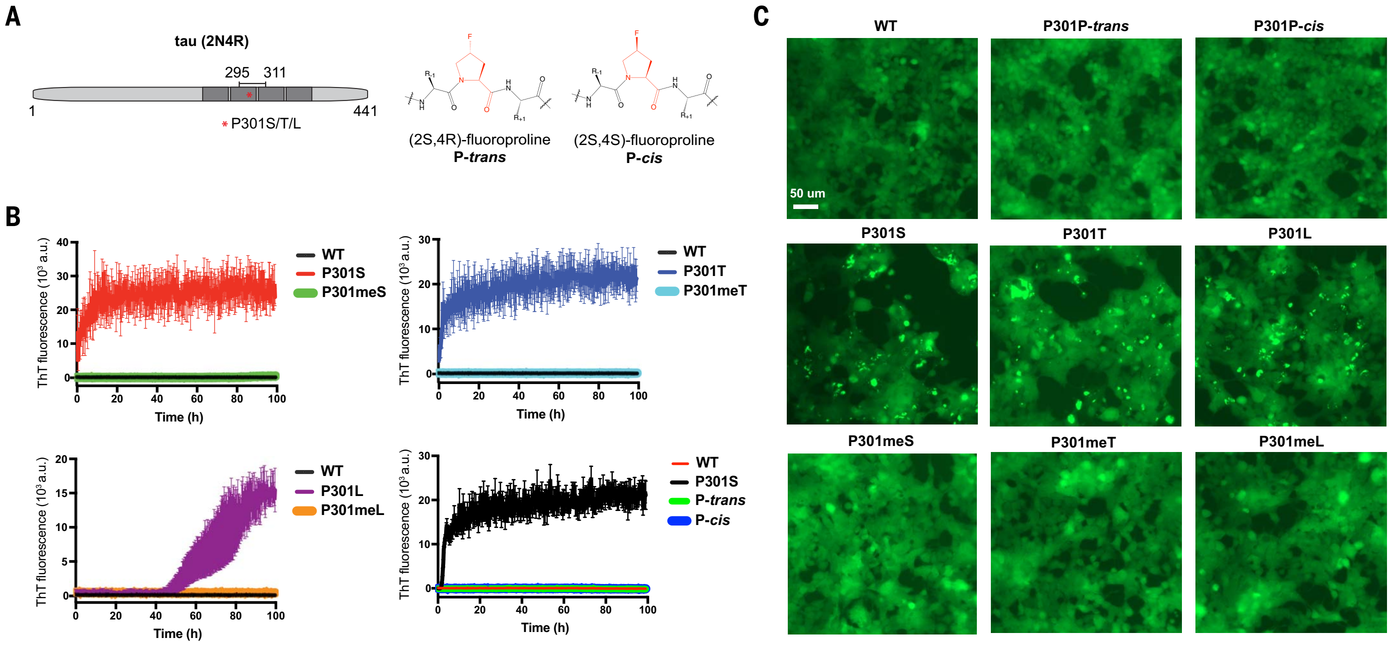


Fig. 8. Restorative effects of eliminating main-chain hydrogen bonding by methyl-capping of peptide backbone nitrogen atoms of the P301S, P301T, and P301L mutational variants of tau. (**A**) Left, schematic of the tau (2N4R) protein with the position of the synthetic tau peptide (residues 295 to 311) and three disease-causing variants (P301S, P301T, and P301L) highlighted with red asterisks. Right, chemical structures of the *trans* (2S,4R; P-*trans*) or *cis* (2S,4S; P-*cis*) conformer of 4-fluoroproline. (**B**) Tau peptides with native sequence of tau residues 295 to 311 (WT), disease-causing variants (P301S, P301T, and P301L), methyl-capped forms of the three disease-causing variants (P301meS, P301meT, and P301meL), and P-*trans* and P-*cis* conformer variants were synthesized and assayed for polymerization as monitored by acquisition of ThT fluorescence. Each of the disease-causing variants led to strong increases in ThT fluorescence relative to WT. No such increase was observed if the peptide backbone nitrogen atoms associated with the disease-variant residues were methyl capped. No enhancement of ThT fluorescence was observed for peptides in which residue P301 was replaced by either the P-*trans* or P-*cis* conformer of 4-fluoroproline. (**C**) Each of the peptides bearing a disease-causing substitution (P301S, P301T, and P301L) led to the formation of distinct aggregates upon assay in the tau biosensor cell line. Whereas WT tau peptide, the peptide nitrogen methyl-capped derivatives of disease-causing variants (P301meS, P301meT, and P301meL), and P-*trans* and P-*cis* conformer variants did not induce endogenous tau to form distinct aggregates in the biosensor cells. Scale bar, 50 μm.

A synthetic peptide carrying dmP at residue 8 was ligated to the remainder of NFL. The purified ligation product was tested for the capacity of the semisynthetic protein to assemble into intermediate filaments. As shown in Fig. 7B, the dmP derivative assembled into intermediate filaments indistinguishable from those formed from the native NFL protein. Consistently, the dmP-containing peptide was equally soluble to the native peptide corresponding to the first 25 residues of the NFL head domain, as assayed by time-dependent acquisition of ThT fluorescence (Fig. 7C). These experiments provide no evidence that restriction of residue P8 of the NFL head domain to its *cis* conformation has any effect on protein function.

The rescue experiments of Fig. 6 strongly suggest that no aspect of residues P8 or P22 of the NFL polypeptide, other than their inability to participate in main-chain hydrogen bonding, is of substantive relevance to function of the NFL head domain. Any of three different amino acids, leucine, glutamine, or arginine, can substitute for P8 so long as their peptide backbone nitrogen atom is methylated. Likewise, arginine, serine, or threonine residues can functionally substitute for P22 so long as their peptide backbone nitrogen atom is methylated. Given the substantive differences between the chemical scaffolds of leucine, glutamine, arginine, serine, and threonine relative to that of proline, we conclude that our observations of functional rescue are likely attributable to removing hydrogen-bonding capacity.

Methylation of the peptide backbone nitrogen of tau mutational variants restores protein solubility

With this insight into why mutation of NFL to several other amino acids in place of P8 or P22 manifests in a biochemical sense, we looked for similarly idiosyncratic proline mutations causative of neurodegenerative disease. Perhaps the most penetrant mutation of the cytoplasmic tau protein causative of neurodegenerative disease maps to residue P301. Disease-causing changes of this P301 residue include substitution by leucine, serine, or threonine (*46*–*50*). Knowing that tau mutations prompt formation of amyloid-like aggregates (*51*, *52*), we reasoned that P301 might insulate against the spread of localized cross-β structure in the normal protein.

Previous studies have already provided compelling evidence that synthetic peptides corresponding to residues 295 to 311 of tau gain a strong propensity to aggregate if residue P301 is changed to either leucine or serine (*52*). The importance of peptide backbone rotation was favored based on the differential effects of substituting proline with either the *cis* (2S,4S) or *trans* (2S,4R) conformer of 4-fluoroproline (Fig. 8A) (*53*). The latter conformer was reported to favor peptide aggregation in a manner similar to the P301L and P301S mutations.

We prepared these same peptides, spanning residues 295 to 311 of the tau polypeptide, and confirmed that the P301L, P301S, and P301T mutations profoundly affected solubility relative to the peptide bearing the native tau sequence, as deduced by ThT fluorescence. We failed, however, to observe any difference in peptide solubility when either the *cis* or *trans* conformer of 4-fluoroproline was used in place of proline at residue 301 (Fig. 8B).

Extending from these observations, we prepared peptide derivatives in which a methyl

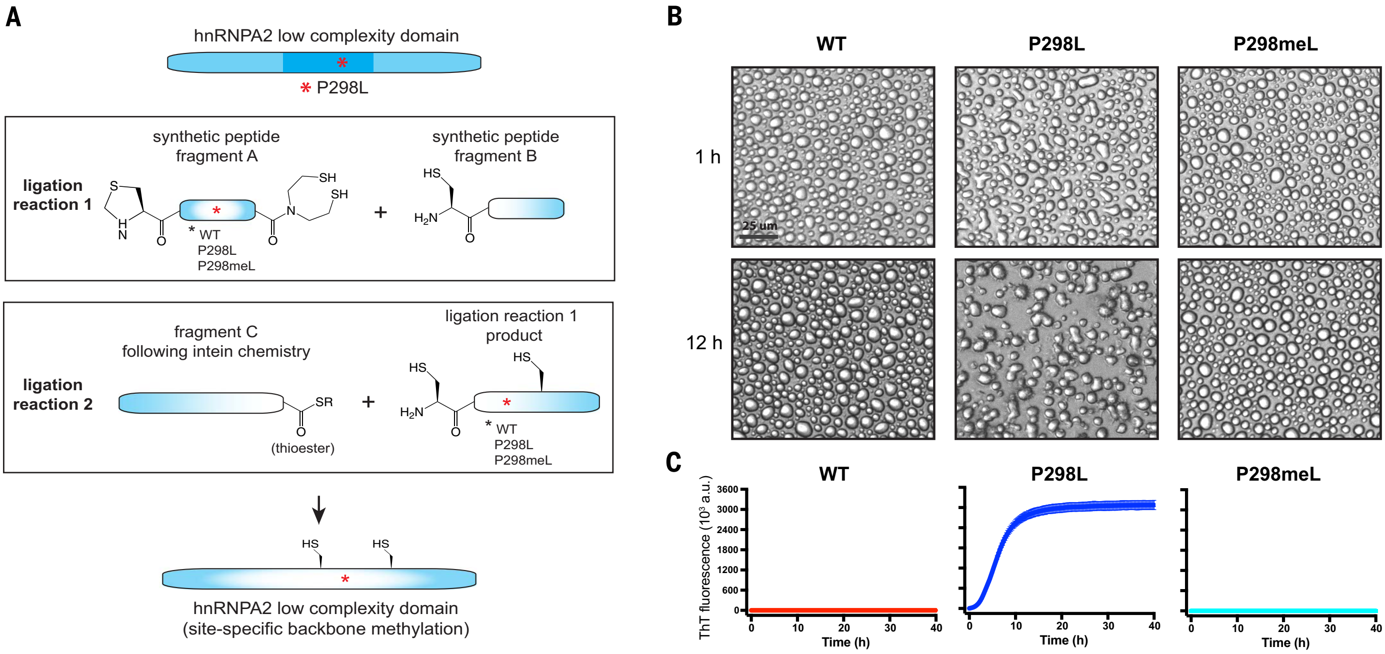


Fig. 9. Restorative effect of eliminating the main-chain hydrogen bond by methyl capping the peptide backbone nitrogen atom of the P298L mutational variant of hnRNPA2. (**A**) Top, schematic representation of the hnRNPA2 LCD with the position of the disease-causing mutant P298L (red asterisk) and self-association core region (dark blue shading) highlighted. Below is an outline of the strategy to prepare semisynthetic variants of the full-length LCD of hnRNPA2. (**B**) The assembled hnRNPA2 LCDs containing the native sequence (WT), P298L variant, and P298meL variant were tested for phase separation in an aqueous buffer. WT hnRNPA2 LCD formed spherical, liquid-like droplets (left). The P298L variant formed distinctly misshapen droplets (middle). The P298meL variant formed spherical droplets indistinguishable from those formed by the native protein (right). Scale bar, 25 μm. (**C**) Synthetic peptides used to assemble semisynthetic proteins assayed in (B) were incubated in the presence of ThT as an assay of time-dependent polymerization. No evidence of fluorescence increase was observed for WT peptide. A large, time-dependent increase in ThT fluorescence was observed for the P298L peptide, and no fluorescence increase was observed for the P298meL peptide.

cap was placed on the peptide backbone nitrogen atom of the leucine, serine, or threonine mutational variant residues. As shown in Fig. 8B, the methyl-capped derivatives of all three tau variant peptides regained solubility indistinguishable from the native tau peptide.

Moving from test tube reactions to living cells, we used a tau biosensor cell line expressing tau-CFP and tau-YFP. The fluorescent tau protein in these cells is homogeneously distributed yet can be converted to a punctate distribution when the cells are exposed to pathological tau aggregates (54, 55). Each of the tau peptides studied in biochemical assays (Fig. 8B) was internalized by lipofection into the tau biosensor cell line. As shown in Fig. 8C, the native tau peptide did not affect the intracellular distribution of fluorescent tau, nor did peptide variants carrying either conformer of 4-fluroproline. Peptides corresponding to the P301L, P301S, and P301T mutations, upon introduction into the tau biosensor cells, led to aggregation of the endogenous tau protein. Finally, synthetic peptide variants of the P301L, P301S, and P301T mutants bearing a methyl cap on the leucine, serine, or threonine residue caused no detectable aggregation of endogenous tau.

Methylation of the peptide backbone nitrogen of an hnRNPA2 mutational variant restores protein function

We also examined the familial, early-onset Paget's disease mutation of residue P298 in the gene encoding hnRNPA2 (56). Our attention to this mutation, aside from its altering of a conserved proline residue, came from the fact that the mutation is located within an LCD long understood to self-associate in a manner leading to phase separation. Indeed, the P298L Paget's disease mutation is located in the middle of the labile cross-β core of the hnRNPA2 protein, as characterized by biochemical footprinting, solid-state NMR spectroscopy, and cryo-EM (56–58).

A three-piece native chemical ligation strategy was implemented to incorporate a synthetic peptide encompassing the site of the P298L mutation into the remainder of the hnRNPA2 LCD (Fig. 9A). Reconstruction of the native hnRNPA2 LCD yielded a protein that phase separated into spherical, liquid-like droplets. The reconstructed variant carrying the P298L mutation formed distinctly misshapen droplets, as has been observed by other investigators (59). Methyl capping of the variant leucine residue yielded a protein that phase separated into spherical, liquid-like droplets indistinguishable from those made by the native hnRNPA2 LCD (Fig. 9B).

To further evaluate variant-mediated enhancement of self-association of the hnRNPA2 LCD, we monitored the time-dependent acquisition of ThT fluorescence. The same 20-residue peptides used to assemble the intact LCD of hnRNPA2 assayed in Fig. 9B were incubated in the presence of ThT. Neither the peptide bearing the native sequence nor that bearing a methyl cap on the leucine residue of the P298L variant exhibited any evidence of polymerization. By contrast, the P298L peptide exhibited robust, time-dependent acquisition of ThT fluorescence (Fig. 9C).

As in the cases of the other nine mutations evaluated in this study, removing the capacity for main-chain hydrogen bonding for the leucine residue of the P298L variant fully restored properly balanced self-associative capacity to the hnRNPA2 LCD. This conclusion can be drawn from assays of both phase separation of the intact LCD (Fig. 9B) and peptide polymerization as visualized by time-dependent acquisition of ThT flourescence (Fig. 9C).

Discussion

We have applied a combination of biochemical and genetic approaches to the study of LCD function. These efforts allowed us to systematically evaluate the importance of 23 peptide backbone amide protons within the TDP-43 LCD for the formation of phase-separated liquid-like droplets. Denial of hydrogen-bonding capacity in a localized region of the TDP-43 LCD distinctly weakened protein self-association. This region of sensitivity to the focused removal of main-chain hydrogen bonds falls within a known cross-β structure.

Our findings comport with the proposed involvement of a small cross-β structure in self-association and phase separation by the TDP-43 LCD, and conform to previous studies of the FUS, hnRNPA2, and ataxin-2 LCDs that also use labile cross-β structures, accounting for the mechanistic basis of phase separation (*9*, *29*, *44*, *60*). These data further agree with studies of the LC head domains functionally essential for assembly of the desmin and NFL intermediate filament proteins (*42*, *43*) and the LC tail domain of the *Drosophila melanogaster* TM1-I/C protein essential for the formation of germ granules located at the posterior tip of fly eggs (*61*).

In addition to studies focused on main-chain hydrogen bonding, glycine-scanning mutagenesis experiments were performed to assess the importance of amino acid side chains to the phenomena of LCD self-association and phase separation. The results confirm the importance of many amino acid side chains, particularly those proximal to the region of the TDP-43 LCD that is both protected from methionine oxidation in the self-associated state and responsible for formation of the cross-β structure.

Whereas the same region of the TDP-43 LCD is sensitive to systematic elimination of the capacity for main-chain hydrogen bonding (Fig. 2 and fig. S1) and systematic variation of side chains (Fig. 4 and fig. S2), close inspection of the data reveals that the different methods of disruption led to different trends. If a variant resulting from peptide backbone methylation had any phenotypic effect on phase separation, then the effect universally weakened liquid-like droplets (Fig. 2 and fig. S1). By contrast, many side chain variants caused the formation of droplets or aggregates that were more urea resistant than those formed by the native TDP-43 LCD.

A second difference between variants made by the two approaches was observed for droplet morphology. Preventing single main-chain hydrogen bonds tended to yield liquid-like droplets with normal spherical morphology. Almost all side chain variants with phenotypic effects either yielded droplets with misshapen morphology or amorphous precipitates, as exemplified by the P320G variant. We suggest that spherically homogeneous droplet morphology may represent a surrogate assay for formation of the properly organized molecular assembly.

Serendipitously, we discovered the importance of residue P320. We propose that by lacking the backbone amide proton essential for main-chain hydrogen bonding, P320 attenuates the network of hydrogen bonds used to form a small and labile cross-β structure located in the ultraconserved region of the TDP-43 LCD. Having noticed β strands on both sides of P320 in cryo-EM structures (Fig. 2F), we propose that restoration of the capacity for main-chain hydrogen bonding at this position might allow for the formation of a long and aberrantly stable cross-β structure. The properties of this aberrant structure are predicted to drive aggregation that restricts TDP-43 to the cytoplasm and kills cultured mammalian cells.

Recognition of the importance of proline 320 to the proper balance of self-association by the TDP-43 LCD led to focused studies of 10 idiosyncratic proline mutations causative of human disease. All 10 mutations localize to one of four proline residues, two in the head domain of NFL, one in tau, and one in hnRNPA2. Functional assays of the NFL, tau, and hnRNPA2 proteins allowed us to investigate how mutations of these four proline residues might manifest at a mechanistic level. In all 10 cases, we were able to functionally repair the mutation-specified variant proteins by eliminating the capacity of the variant amino acid residue to participate in main-chain hydrogen bonding.

These studies revealed functional rescue of the NFL P8L variant by either methylation of the peptide backbone nitrogen of the variant leucine residue or replacement of this nitrogen with oxygen. Amide nitrogen methylation can influence *phi:psi* angles of the peptide backbone (*62*). It is likewise possible that the amide-to-ester chemical change might influence protein function in some manner other than elimination of the capacity for hydrogen bonding. That both backbone nitrogen methylation and esterification resulted in full correction of protein function favors the parsimonious conclusion that both chemical modifications manifest restorative activity by eliminating the capacity for hydrogen bonding.

The results reported herein help to bridge human genetics and biochemistry. If our interpretations are correct, then we are now able to understand how the variant proteins encoded by certain proline mutations change the properties of the LCD in which they reside. Our experiments should also facilitate bioinformatic searches for similarly idiosyncratic proline mutations described by geneticists over the past several decades. We predict that if a proline residue is the locus of recurrent mutations, if the mutations are dominant, and if the mutations lie within an LCD, then the mechanistic basis for their deleterious effects might match what we describe for mutations in NFL, tau, and hnRNPA2.

Here, we began with biochemical experiments and leveraged the knowledge gained to examine disease mutations identified by human genetics. Overall, our results yield a common interpretation. Protein domains of low sequence complexity self-associate through the formation of labile structures poised at the threshold of equilibrium. Despite their labile nature, there is sequence specificity to the formation of these structures. This specificity is dictated by the cardinal rules of structural biology taught to us by Pauling, Perutz, Kendrew, and the icons of x-ray crystallography. Our interpretations give evidence that evolution of LCDs has favored the formation of protein chains designed to self-associate weakly. Deleterious effects result if these domains bear mutation-specified variation causative of enhanced self-association.

These and previous studies have mapped the regions responsible for LCD self-association to relatively small cross-β cores constituting no more than a fraction of the intact LCD. Once these localized cross-β cores self-associate, the remaining regions of the LCD dangle off in an unstructured state. Others have suggested how these unstructured regions that flank the transiently formed cross-β elements of LCDs might be of biological utility. Tyrosine, phenylalanine, or methionine residues might function as adhesive "stickers" to facilitate nonspecific interactions with other LCDs (*63*), thus allowing the self-assembled cross-β core to nucleate a larger and more complex assembly. It is likewise possible that small sequence elements designated LARKS (low-complexity aromatic-rich kinked segments) may facilitate biologically useful tertiary interactions (*64*).

The disease-associated mutational variants of the NFL head domain and the hnRNPA2 LCD studied herein localize to cross-β cores, not to the unstructured protein regions located outside of the cores. Core-proximal ALS-causing mutations have also been characterized for FUS, hnRNPA1, hnRNPDL, and a mutational variant of hnRNPA2 different from that characterized in this study (*57*, *65*). A polypeptide region that exists in conformational equilibrium between a folded and unfolded state could have the folded state stabilized by a single missense mutation. By contrast, we predict that the opportunity for mutational variants to influence LCD function is diminished if they occur in polypeptide regions that do not specify labile, structure-based self-association.

All cross-β interactions allowing for LCD self-association reported to date are organized in a parallel, "in-register" conformation (*27*, *29*, *32*, *61*, *66*). This organization aligns each residue along the self-interacting interface with its sister residue of the paired polypeptide chain. Despite the fact that hydrogen

bonding along the polypeptide main chain contributes significant interaction energy to each structure, amino acid side chains are undoubtedly important both for interaction specificity and properly tuned avidity. Given the small sizes of the cross-β–forming regions, coupled with the uniform, in-register geometry of pairing, it may someday be possible to deduce an interaction code that will allow for computational identification of the regions of LCDs responsible for self-interaction.

We close with two considerations. First, the labile interactions conceptualized in this and other studies of LCD self-association likely assist in thousands of different cellular processes. Second, the labile and reversible nature of these interactions frequently leaves LCDs in an unstructured state readily accessible for almost any form of posttranslational modification. The combination of these features raises the possibility that many forms of dynamic cell behavior may be controlled by regulatory events directly operative upon LCDs.

Materials and Methods

Molecular cloning

Standard gene cloning and mutagenesis strategies were used to generate expression constructs for all recombinant proteins.

For plasmid constructs used in the glycine mutagenesis scanning experiment, wild-type TDP-43 LCD (residues 262 to 414) and its variants were cloned into pHis-parallel vector to yield His-TDP-43 LCD constructs. Recombinant GyrA intein fusion fragments of the TDP-43 LCD that were required for expressed protein ligation were cloned into the pHis-parallel vector. The complete scan required two distinct three-piece ligation strategies and thus two separate intein fusion fragments (strategy A, piece 1, residues 262 to 314; strategy b, piece 1, residues 262 to 326) of the TDP-43 LCD. The two separate, recombinant piece three constructs (strategy A, piece 3, residues 328 to 414; strategy B, piece 3, residues 341 to 414) of the TDP-43 LCD were fused with a 6×His-SUMO tag at the N terminus and a GryA intein at the C terminus in pET-28a vector. Wild type (WT) TDP-43 and mutant constructs for transient transfection of cultured U2OS cells were cloned as N-terminal GFP fusions in the pCDNA 3.1 vector with an N-terminal FLAG tag. For stable, inducible TDP-43 expression in mammalian cells, full-length WT, P320G, and M337V were constructed in the CMV mammalian expression vector containing the Tet operator and an N-terminal FLAG-GFP tag.

For native chemical ligation-compatible NFL fragments, an N-terminal 6×His-GFP tag was fused to constructs that would ultimately yield the head domain (residues 27 to 87) and full-length (residues 27 to 543) constructs. A caspase 3 cleavage site (GDEVDC) was inserted between the GFP and NFL fragments such that digestion would yield NFL fragments bearing an N-terminal cysteine at amino acid position 27. These constructs were built into the pHis-parallel vector.

The hnRNPA2-intein fusion fragment (hnRNPA2 residues 181 to 284 with a C-terminal GyrA intein fusion) was inserted in the pHis-parallel vector. A 6×His tag was appended to the C terminus, and the N-terminal 6×His tag was omitted during cloning. The N-terminal 6×His-tagged WT hnRNPA2 LCD (residues 181 to 341) and its variants were made by inserting the LCD fragment into the pHis-parallel vector.

Recombinant TDP-43 piece 1 and piece 3 sequences

Strategy 1, piece 1: 6xHis-TDP-43 (residues 262 to 314)SYYHHHHHHDYDIPTTENLYFQGAMDPEF*PKHNSNRQLERSGRFGGNPGGFGNQGGFGNSRGGGAGLGNNQGSNMGGGMNFG*-thioesterStrategy 2, piece 1: 6xHis-TDP-43 (residues 262 to 327)SYYHHHHHHDYDIPTTENLYFQGAMDPEF*PKHNSNRQLERSGRFGGNPGGFGNQGGFGNSRGGGAGLGNNQGSNMGGGMNFGAFSINPAMMAAAQA*-thioesterStrategy 1, piece 3: TDP-43 (residues 329 to 414; A329C)CLQSSWGMMGMLASQQNQSGPSGNNQNQGNMQREPNQAFGSGNNSYSGSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSSGWGMStrategy 2, piece 3: TDP-43 (residues 341 to 414; A341C)CSQQNQSGPSGNNQNQGNMQREPNQAFGSGNNSYSGSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSSGWGM.

Recombinant NFL and hnRNPA2 fragment sequences

hnRNPA2 (residues: 181 to 284)MQEVQSSRSGRGGNFGFGDSRGGGGNFGPGPGSNFRGGSDGYGSGRGFGDGYNGYGGGPGGGNFGGSPGYGGGRGGYGGGGPGYGNQGGGYGGGYDNYGGGNYG-thioesterNFL (S27C, residues 27 to 86, for head domain assembly)CSVRSGYSTARSAYSSYSAPVSSSLSVRRSYSSSSGSLMPSLENLDLSQVAAISNDLKSINFL (S27C, residues 27 to 543, for full length assembly)CSVRSGYSTARSAYSSYSAPVSSSLSVRRSYSSSSGSLMPSLENLDLSQVAAISNDLKSIRTQEKAQLQDLNDRFASFIERVHELEQQNKVLEAELLVLRQKHSEPSRFRALYEQEIRDLRLAAEDATNEKQALQGEREGLEETLRNLQARYEEEVLSREDAEGRLMEARKGADEAALARAELEKRIDSLMDEISFLKKVHEEEIAELQAQIQYAQISVEMDVTKPDLSAALKDIRAQYEKLAAKNMQNAEEWFKSRFTVLTESAAKNTDAVRAAKDEVSESRRLLKAKTLEIEACRGMNEALEKQLQELEDKQNADISAMQDTINKLENELRTTKSEMARYLKEYQDLLNVKMALDIEIAAYRKLLEGEETRLSFTSVGSITSGYSQSSQVFGRSAYGGLQTSSYLMSTRSFPSYYTSHVQEEQIEVEETIEAAKAEEAKDEPPSEGEAEEEEKDKEEAEEEEAAEEEEAAKEESEEAKEEEEGGEGEEGEETKEAEEEEKKVEGAGEEQAAKKKD.

Recombinant protein expression and purification

All recombinant proteins used in this study were expressed in *Escherichia coli* BL21 (DE3) cells grown to an optical density at 600 nm of 0.6 in LB medium.

Expression of His-TDP-43 LCD proteins and piece 1 proteins were induced with 0.8 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37°C for 4 hours. Expression of piece 3 proteins was induced with 0.6 mM IPTG at 16°C for 16 hours. The cells were harvested by centrifugation at 4000*g* for 15 min.

Expression of NFL proteins was induced with 0.6 mM IPTG at 16°C for 16 hours. Cells were harvested by centrifugation at 4000*g* for 15 min. The hnRNPA2 LCD proteins were expressed by the addition of 0.8 IPTG at 37°C, and cells were harvested after 4 hours.

For purification of His-TDP-43 LCD WT and its variants, the cell pellets were resuspended in buffer A containing 25 mM Tris-HCl (pH 7.5), 200 mM NaCl, 6 M guanidine-HCl, 10 mM β-mercaptoethanol (β-ME), and 20 mM imidazole, and then disrupted by sonication. The cell lysates were clarified by centrifugation at 40,000*g* for 50 min. Supernatant was then applied to Ni^{2+}-NTA resin (Qiagen), and the column washed with buffer A and eluted with buffer A supplemented with 300 mM imidazole. Pure protein was concentrated by centrifugal filtration (Amicon Ultra-15) and samples were aliquoted, flash frozen, and stored at –80°C for future use.

For purification of recombinant TDP-43 LCD piece 1 protein thioesters, the cell pellets were resuspended in buffer B, containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl, 8 M urea, 1 mM tris(2-carboxyethyl) phosphine (TCEP), and 20 mM imidazole, and then disrupted by sonication. The resulting cell lysates were centrifuged at 40,000*g* for 50 min. The supernatants were applied to Ni^{2+}-NTA resin (Qiagen), and the column was washed with buffer B and eluted with buffer B supplemented with 300 mM imidazole. The purified protein was refolded by overnight dialysis against a buffer containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl, 1 M urea, and 1 mM TCEP. The refolded protein was centrifugated at 4000*g* for 30 min to remove precipitation, and the supernatant was supplemented with 300 mM MES-Na, and 10 mM TCEP, adjusted to pH 7.0, and incubated at room temperature for 16 hours to achieve thiolysis. After thiolysis, the protein solution was centrifuged at 4000*g* for 30 min to isolate precipitation. Precipitated protein was resolubilized in a buffer containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl,

6 M guanidine-HCl, and 1 mM TCEP, and then combined with the supernatant fraction and loaded onto Ni^{2+}-NTA resin. The column was washed with a buffer containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl, 6 M guanidine-HCl, 1 mM TCEP, and 20 mM imidazole to remove unbound proteins. Piece 1 thioester was then eluted with the same buffer supplemented with 300 mM imidazole. The eluted piece 1 fragment was supplemented with 1% trifluoroacetic acid (TFA) and purified by reverse-phase (RP)-HPLC (C4 preparative column). Fractions were analyzed by RP-HPLC and intact electrospray ionization mass spectrometry (ESI-MS), and pure fractions were lyophilized.

For purification of recombinant TDP-43 LCD piece 3 constructs, the cell pellets were resuspended in buffer C containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl, 2 M urea, 1 mM TCEP, and 20 mM imidazole, and then disrupted by sonication. The cell lysates were clarified by centrifugation at 40,000*g* for 50 min, supernatants were applied to Ni^{2+}-NTA resin, and the column was washed with buffer C and eluted with buffer C supplemented with 300 mM imidazole. The eluted proteins were dialyzed against buffer C without imidazole for 3 hours to remove imidazole. After dialysis, Ulp1 enzyme was added to the protein solution for 1 hour to remove the SUMO tag, and 200 mM β-ME was added for 16 hours to achieve GryA fusion hydrolysis. The solution was passed over Ni^{2+}-NTA resin to capture cleaved SUMO and GryA proteins. Flow-through containing piece 3 was collected, supplemented with 1% TFA, and purified by HPLC (C4 preparative column). Fractions were analyzed by RP-HPLC, and intact ESI-MS and pure fractions were lyophilized.

For purification of NFL fragments, the cell pellets were resuspended in a lysis buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 M urea, 5 mM β-ME, and 20 mM imidazole, and then disrupted by sonication. The cell lysates were clarified by centrifugation at 36,000*g* for 50 min, supernatants were applied to Ni^{2+}-NTA resin (Qiagen), and the column was washed with the lysis buffer and eluted with lysis buffer supplemented with 300 mM imidazole. The eluted proteins were dialyzed against lysis buffer without imidazole for 3 hours. The caspase-3 enzyme was then added to the protein solution for 2 hours to cleave the GFP tag while generating the N-terminal cysteine. The protein solution was again passed over Ni^{2+}-NTA resin to capture the GFP fusion tag and any undigested protein. Flow-through containing NFL fragments was collected, supplemented with 1% TFA, and purified by RP-HPLC (C4 preparative column). Purity was determined by analytical RP-HPLC and ESI-MS, and pure fractions were pooled and lyophilized.

For purification of the hnRNPA2 LCD fragment, the cell pellet was resuspended in a lysis buffer containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl, 8 M urea, 1 mM TCEP, and 20 mM imidazole, and then disrupted by sonication. The resulting cell lysates were centrifugated at 40,000*g* for 50 min. The supernatants were applied to Ni^{2+}-NTA resin, and the column was washed with lysis buffer and eluted with lysis buffer supplemented with 300 mM imidazole. The purified protein was refolded by overnight dialysis against a buffer containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl, 1 M urea, and 1 mM TCEP. The refolded protein was then centrifugated at 4000*g* for 30 min to remove any precipitates, and the cleared supernatant was supplemented with 300 mM MES-Na and incubated at room temperature for 16 hours. After the GryA thiolysis reaction, the protein solution was centrifuged at 4000*g* for 30 min to pellet any precipitate. The precipitated protein was resolubilized by a buffer containing 25 mM potassium phosphate (pH 7.2), 150 mM NaCl, 6 M guanidine-HCl, and 1 mM TCEP, and then combined with the supernatant fraction and loaded on Ni^{2+}-NTA resin. The flow-through containing free hnRNPA2 LCD fragment was collected, supplemented with 1% TFA, and purified by RP-HPLC. Purity was determined by analytical RP-HPLC and ESI-MS, and pure fractions were pooled and lyophilized.

Preparation of SEA resin

Bis(2-sulfanylethyl)amido (SEA) polystyrene resin was made following the protocol from Ollivier *et al.* (67).Piece 2 TDP-43 peptides with C-terminal SEA (strategy 1, residues 315 to 328; strategy 2, residues 328 to 340).

Strategy 1, piece 2 peptides

Sequence of TDP-43 WT (residues 315 to 328, A315Thz)-SEA: Thz-FSINPAMMAAAQA-SEASequence of TDP-43 F316meF (residues 315 to 328, A315Thz, F316meF)-SEA: Thz-meF-SINPAMMAAAQA-SEASequence of TDP-43 S317meS (residues 315 to 328, A315Thz, S317MeS)-SEA: Thz-FmeSINPAMMAAAQA-SEASequence of TDP-43 I318meI (residues 315 to 328, A315Thz, I318MeI)-SEA: Thz-FSmeINPAMMAAAQA-SEASequence of TDP-43 N319meN (residues 315 to 328, A315Thz, N319meN)-SEA: Thz-FSImeNPAMMAAAQA-SEASequence of TDP-43 A321meA (residues 315 to 328, A315Thz, A321meA)-SEA: Thz-FSINPmeAMMAAAQA-SEASequence of TDP-43 M322meM (residues 315 to 328, A315Thz, M322meM)-SEA: Thz-FSINPAmeMMAAAQA-SEASequence of TDP-43 M323meM (residues 315 to 328, A315Thz, M323meM)-SEA: Thz-FSINPAMmeMAAAQA-SEASequence of TDP-43 A324meA (residues 315 to 328, A315Thz, A324meA)-SEA: Thz-FSINPAMMmeAAAQA-SEASequence of TDP-43 A325meA (residues 315 to 328, A315Thz, A325meA)-SEA: Thz-FSINPAMMAmeAAQA-SEASequence of TDP-43 A326meA (residues 315 to 328, A315Thz, A326meA)-SEA: Thz-FSINPAMMAAmeAQA-SEASequence of TDP-43 Q327meQ (residues 315 to 328, A315Thz, Q327meQ)-SEA: Thz-FSINPAMMAAAmeQA-SEASequence of TDP-43 A328meA (residues 315 to 328, A315Thz, A328meA)-SEA: Thz-FSINPAMMAAAQmeA-SEASequence of TDP-43 A321meA-M323meM-A325meA (residues 315 to 328, A315Thz, A321meA-M323meM-A325meA)-SEA: Thz-FSINPmeAMmeMAmeAAQA-SEASequence of TDP-43 M323meM-A325meA-Q327meQ (residues 315 to 328, A315Thz, M323meM-A325meA-Q327meQ)-SEA: Thz-FSINPAM-meMAmeAAmeQA-SEA.

Strategy 2, piece 2 peptides

Sequence of TDP-43 WT (residues 328 to 340, A328Thz)-SEA: Thz-ALQSSWGMMGML-SEA-Sequence of TDP-43 A329meA (residues 328 to 340, A329meA, A328Thz)-SEA Thz-meALQSSWGMMGML-SEASequence of TDP-43 L330meL (residues 328 to 340, L330meL, A328Thz)-SEA Thz-AmeLQSSWGMMGML-SEASequence of TDP-43 Q331meQ (residues 328 to 340, Q331meQ, A328Thz)-SEA Thz-ALmeQSSWGMMGML-SEASequence of TDP-43 S332meS (residues 328 to 340, S332meS, A328Thz)-SEA Thz-ALQmeSSWGMMGML-SEASequence of TDP-43 S333meS (residues 328 to 340, S333meS, A328Thz)-SEA Thz-ALQSmeSWGMMGML-SEASequence of TDP-43 W334meW (residues 328 to 340, W334meW, A328Thz)-SEA Thz-ALQSSmeWGMMGML-SEASequence of TDP-43 G335meG (residues 328 to 340, G335meG, A328Thz)-SEA Thz-ALQSSWmeGMMGML-SEASequence of TDP-43 M336meM (residues 328 to 340, M336meM, A328Thz)-SEA Thz-ALQSSWGmeMMGML-SEASequence of TDP-43 M337meM (residues 328 to 340, M337meM, A328Thz)-SEA Thz-ALQSSWGMmeMGML-SEASequence of TDP-43 G338meG (residues 328 to 340, G338meG, A328Thz)-SEA Thz-ALQSSWGMMmeGML-SEASequence of TDP-43 M339meM (residues 328 to 340, M339meM, A328Thz)-SEA Thz-ALQSSWGMMGmeML-SEA.

NFL peptides with C-terminal SEA

NFL WT (residues 2 to 26, I26L): SSFSYE-PYYSTSYKRRYVETPRVHL-SEANFL P8L (residues 2 to 26, P8L, I26L): SSFSYELYYSTSY-KRRYVETPRVHL-SEANFL P8meL (residues 2 to 26, P8meL, I26L): SSFSYEmeLYYSTSYKR-RYVETPRVHL-SEANFL P8Q (residues 2 to 26, P8Q, I26L): SSFSYEQYYSTSYKRRYVET-PRVHL-SEANFL P8meQ (residues 2 to 26, P8meQ, I26L): SSFSYEmeQYYSTSYKRRY-VETPRVHL-SEANFL P8R (residues 2 to 26, P8R, I26L): SSFSYERYYSTSYKRRYVETPRVHL-SEANFL P8meR (residues 2 to 26, P8meR,

I26L): SSFSYEmeRYYSTSYKRRYVETPRVHL-SEANFL P8dmP (residues 2 to 26, P8dmP, I26L): SSFSYEdmPYYSTSYKRRYVETPRVHL-SEANFL P8esL (residues 2 to 26, P8esL, I26L): SSFSYEesLYYSTSYKRRYVETPRVHL-SEANFL P22R (residues 2 to 26, P22R, I26L): SSFSYEPYYSTSYKRRYVETRRVHL-SEANFL P22meR (residues 2 to 26, P22meR, I26L): SSFSYEPYYSTSYKRRYVETmeRRVHL-SEANFL P22S (residues 2 to 26, P22S, I26L): SSFSYEPYYSTSYKRRYVETSRVHL-SEANFL P22meS (residues 2 to 26, P22meS, I26L): SSFSYEPYYSTSYKRRYVETmeSRVHL-SEANFL P22T (residues 2 to 26, P22T, I26L): SSFSYEPYYSTSYKRRYVETTRVHL-SEANFL P22meT (residues 2 to 26, P22meT, I26L): SSFSYEPYYSTSYKRRYVETmeTRVHL-SEA.

hnRNPA2 peptides with C-terminal SEA

hnRNP A2 WT (residues 285 to 305, A285Thz): Thz-GNYNDFGNYNQQPSNYGPMK-SEAhnRNP A2 P298L (residues 285 to 305, P298L, A285Thz): Thz-GNYNDFGNYNQQLSNYGPMK-SEAhnRNP A2 P298meL (residues 285 to 305, P298meP, A285Thz): Thz-GNYNDFGNYNQQmeLSNYGPMK-SEAwhere, SEA = bis(2-sulfanylethyl)amido group, dmP = L-5,5-dimethylproline, me = N$^{\alpha}$-methyl, es = ester, Thz = thiazolidine.

Amidated peptides described in this study

hnRNPA2 (306-341, S306C):CGNFGGSRNMG-GPYGGGNYGPGGSGGSGGYGGRSRY-CONH$_2$tau-RD (residues 295 to 311): DNIKHVPGGGSVQIVYK-CONH$_2$tau-RD P301S (residues 295 to 311, P301S): DNIKHVSGGGSVQIVYK-CONH$_2$tau-RD P301meS (residues 295 to 311, P301meS): DNIKHV-meSGGGSVQIVYK-CONH$_2$tau-RD P301T (residues 295 to 311, P301T): DNIKHVTGGGSVQIVYK-CONH$_2$tau-RD P301meT (residues 295 to 311, P301meT): DNIKHVmeTGGGSVQIVYK-CONH$_2$tau-RD P301L (residues 295 to 311, P301L): DNIKHVLGGGSVQIVYK-CONH$_2$tau-RD P301meL (residues 295 to 311, P301meL): DNIKHV-meLGGGSVQIVYK-CONH$_2$tau-RD P301P-*cis* (residues 295 to 311, P301P-*cis*): DNIKHVP-*cis*GGGSVQIVYK-CONH$_2$tau-RD P301P-*trans* (residues 295 to 311, P301P-*trans*) DNIKHVP-*trans*GGGSVQIVYK-CONH$_2$where me = N$^{\alpha}$-methyl.

Peptide synthesis

All fluorenylmethyloxycarbonyl (Fmoc)–protected amino acids were purchased from Oakwood Chemical or Combi-Blocks. Peptide synthesis resins (Trityl-OH ChemMatrix and Rink Amide ChemMatrix) were purchased from Biotage. All analytical RP-HPLC was performed on an Agilent 1260 series instrument equipped with a quaternary pump and an XBridge Peptide C18 or C4 column (5 μm, 4 × 150 mm, Waters) at a flow rate of 1 ml/min. Similarly, semipreparative scale purifications were performed with a XBridge Peptide C18 or C4 semipreparative column (5 μm, 10 mm × 250 mm, Waters) at a flow rate of 4 ml/min. Preparative RP-HPLC was performed on an Agilent 1260 series instrument equipped with a preparatory pump and a XBridge Peptide C18 or C4 preparatory column (10 μM; 19 × 250 mm, Waters) at a flow rate of 20 ml/min. All instruments were equipped with a variable wavelength UV detector. All RP–HPLC steps were performed using 0.1% TFA (Oakwood Chemical) in H_2O (solvent A) and 90% acetonitrile (Sigma-Aldrich) or 0.1% TFA in H_2O (solvent B) as mobile phases. For liquid chromatography–mass spectrometry (LC-MS) analysis, 0.1% formic acid (Sigma-Aldrich) was substituted for TFA in mobile phases. Mass analysis was performed for each product on an LC-MS (Agilent Technologies) equipped with a 300SB-C18 column (3.5 μM; 4.6 × 100 mm, Agilent Technologies) or a X500B QTOF (Sciex).

All peptides containing C-terminal SEA were synthesized by solid-phase peptide synthesis on a CEM Discover Microwave Peptide Synthesizer (Matthews, NC) using the Fmoc-protection strategy on SEA resin (0.16 mmol/g, Iris Biotech). SEA resin was washed and swelled with *N,N* dimethylformamide (DMF, Oakwood Chemical) and bubbled in N_2 for 15 min. For manual loading of the first amino acid to SEA resin, Fmoc-alanine (10 eq; for strategy 1 peptides), Fmoc-leucine (10 eq; for strategy 2 peptides and NFL peptides), Fmoc-lysine (10 eq; hnRNPA2 peptides), HATU (10eq, Oakwood Chemical), and DIPEA (30eq, Sigma-Aldrich) were mixed in DMF, and the resin was bubbled with N_2 for 1 hour. This step was repeated with fresh reagents to ensure complete loading. The resin was then washed with DMF and bubbled in acetic anhydride:DIPEA (20 eq:40 eq) in DMF for 20 min to quench unreacted sites. Subsequent peptide synthesis reactions were performed on an automated microwave synthesizer. For coupling reactions, amino acids (5 eq) were activated with *N,N*-diisopropylcarbodiimide (DIC) (5 eq, Oakwood Chemical):Oxyma (5 eq, Oakwood Chemical) and heated to 90°C for 2 min while bubbling with nitrogen gas in DMF. Fmoc deprotection was performed with 20% piperidine (Sigma-Aldrich) in DMF supplemented with 0.1 M 1-hydroxybenzotriazole hydrate (HOBt) (Oakwood Chemical) at 90°C for 1 min while bubbling with nitrogen gas. Resin cleavage was performed with 95% TFA, 2.5% triisopropylsilane (TIS, Sigma-Aldrich), and 2.5% H_2O for 3 h at 25°C. The crude peptide was then precipitated by the addition of a 10-fold volume of ice-cold ether and centrifuged at 4000 relative centrifugal force (RCF) for 10 min at 4°C. To oxidize the C-terminal SEA, the pellet was washed with ice-cold ether and resuspended with buffer containing 6 M guanidine-HCl, 0.1 M sodium phosphate, and 5% (v/v) dimethyl sulfoxide at pH 7.5 to 8.0. This reaction was nutated at 25°C for 16 hours, and SEA ring oxidation was confirmed by a −2 Da mass change using LC-MS. To deprotect the Thz ring, 200 mM of O-methylhydroxylamine HCl (Combi-Blocks) was added to the reaction mixture, the pH of the reaction was adjusted to 4.0, and the mixture was incubated at 37°C for 1 hour. Deprotection of the thiazolidine ring was confirmed by a −12 Da mass change using LC-MS. Oxidized and Thz-deprotected peptides were purified by preparative C4 RP-HPLC. Fractions were analyzed on analytical C4 RP-HPLC and ESI-MS, and those containing pure product (>95%) were pooled, lyophilized, and stored at −80°C.

The above amidated peptides were synthesized by solid-phase peptide synthesis on a CEM Discover Microwave Peptide Synthesizer (Matthews, NC) using the Fmoc-protection strategy on Rink Amide-ChemMatrix resin (0.5 mmol/g). For coupling reactions, amino acids (5 eq) were activated with DIC (5 eq, Oakwood Chemical)/Oxyma (5 eq, Oakwood Chemical) and heated to 90°C for 2 min while bubbling with N_2 in DMF (Oakwood Chemical). Fmoc deprotection was performed with 20% piperidine (Sigma-Aldrich) in DMF supplemented with 0.1 M HOBt (Oakwood Chemical) at 90°C for 1 min while bubbling with N_2. Cleavage from resin was performed with 92.5% TFA, 2.5% TIS (Sigma-Aldrich), 2.5% 1,2-ethanedithiol (Sigma-Aldrich), and 2.5% H_2O for 2 h at 25°C on an end-over-end rotisserie. The crude peptide was then precipitated by the addition of a 10-fold volume of ice-cold ether and centrifuged at 4000 RCF for 10 min at 4°C. The pellet was resuspended in solvent A and purified by preparative C18 RP-HPLC. Fractions were analyzed on analytical C18 RP-HPLC and ESI-MS, and those containing pure product (>95%) were pooled, lyophilized, and stored at −80°C.

For peptide chain elongation from the dmP building block, resin was removed from the synthesizer and manual coupling was performed with Fmoc-5,5-dmP-OH (6 eq, Iris Biotech), PyAOP (6 eq, Oakwood Chemical), and DIPEA (12 eq, Sigma-Aldrich) for 1 hour at 50°C. Coupling was repeated with fresh reagents to maximize yield. Resin was acetyl capped (acetic anhydride:DIPEA, 20 eq:40 eq in DMF) to quench unreacted sites and placed back on an automated synthesizer to complete synthesis.

The L-leucic acid building block was manually coupled to resin in a reaction containing L-leucic acid (5 eq, Combi-Blocks), DIC (5 eq), and oxyma (5 eq) in DMF for 30 min at 20°C. Coupling was repeated twice with fresh reagents to maximize yield. For peptide chain elongation from the L-leucic acid building block, a reaction containing amino acid [5 eq, Fmoc-Glu(OtBu)-OH], MSNT (5 eq, Sigma-Aldrich), and 1-methylimidazole (3.75 eq, Sigma-Aldrich) in dichloromethane was purged with N_2 and incubated for 2 hours on an end-over-end rotisserie. Coupling was repeated three

times with fresh reagents to maximize yield. After the final reaction, resin was placed back on an automated synthesizer to complete synthesis.

Assembly of semisynthetic TDP-43 constructs

To scan N^{α}-methyl amino acids from amino acids 316 to 328, all strategy 1 pieces described above were assembled. To scan N^{α}-methyl amino acids from amino acids 329 to 339, all strategy 2 pieces described above were assembled. Two separate strategies were required because we were unable to synthesize a single peptide from residues 316 to 339 on the requisite SEA resin at sufficient yields. The following three-piece assembly strategy was performed for both the strategy 1 and strategy 2 assemblies.

Piece 1 (bearing MES thioester) and piece 2 (bearing free N-terminal cysteine and an oxidized SEA ring) were combined at ~10 mM piece 1 and ~8 mM piece 2 in a degassed buffer of 6 M guanidine-HCl, 0.1 M sodium phosphate, and 100 mM 2,2,2-trifluoroethanethiol (TFET, Sigma-Aldrich) adjusted to pH 7.0. Excess piece 2 was used to push all peptide to ligated product and ensure efficient separation of ligated product from starting material during purification. Reactions were incubated at 37°C for 4 hours and progress was monitored by C4 RP-HPLC and ESI-MS analysis. The C-terminal oxidized SEA ring remained intact and inert throughout this reaction. Ligation products were purified on a semipreparative C4 RP-HPLC column and fractions containing target mass were pooled, lyophilized, and stored at –80°C. This product is referred to as "piece 1+2."

The lyophilized piece 1+2 construct was resuspended in 0.3 to 1 ml of degassed buffer (final piece 1+2 concentration ~0.5 mM) containing 6 M guanidine-HCl, 0.1 M sodium phosphate, 200 mM MES-Na (Sigma-Aldrich), and 50 mM TCEP, and then adjusted to pH 4.0. This reaction was incubated at 37°C for 16 hours to convert the C-terminal SEA moiety to a MES-thioester as confirmed by a –5 Da mass shift using LC-MS. Once MES conversion was complete, piece 3 (bearing an N-terminal cysteine) was directly added to this reaction mixture to a concentration of ~1.5 mM. Excess piece 3 was used to push all piece 1+2 to ligated product and ensure efficient separation of ligated product from starting material during purification. Reactions were supplemented with 20 mM TCEP and 100 mM TFET, adjusted to pH 7.0, and incubated 37°C for 16 hours. Upon completion of the ligation reaction as judged by RP-HPLC and ESI-MS analysis, excess TFET was removed by dialyzing the reaction mixture against a buffer containing 6 M guanidine-HCl and 0.1 M sodium phosphate at pH 4.0 for 3 hours. Desulfurization of the final product to the native TDP-43 LCD sequence achieved through free radical–mediated desulfurization by addition of 200 mM TCEP, 60 mM VA-044 radical initiator, and 150 mM of reduced glutathione. Reactions were adjusted to pH 7.0 and incubated at 37°C for 16 hours. Ligation products were purified on a semipreparative C4 RP-HPLC column, and pure fractions were pooled, lyophilized, and stored at –80°C. Final purities of >95% were judged by analytical RP-HPLC, ESI-MS analysis, and SDS-polyacrylamide gel electrophoresis (PAGE).

Assembly of semisynthetic NFL constructs

To assemble NFL head domains (residues 2 to 86), native chemical ligation reactions were performed by combining 5 mM peptide bearing C-terminal SEA moieties (NFL residues residues 2 to 26; wt, P8Q, PmeQ, P8dmP, P22S or P2meS) with 1 mM of the NFL head domain fragment bearing an N-terminal cysteine (S27C, residues 27-86) in a degassed buffer of 6 M guanidine-HCl, 0.1 M sodium phosphate, 20 mM TCEP, 200 mM MES-Na (Sigma-Aldrich), and 150 mM 2,2,2-trifluoroethanethiol (TFET, Sigma-Aldrich). Reaction pH was adjusted to 7.0, and the mixture was incubated at 37°C for 16 hours. Reaction progress was monitored by RP-HPLC and ESI-MS analysis. Ligation products were purified on a preparative C18 RP-HPLC column and fractions were analyzed by RP-HPLC and ESI-MS. Fractions containing target mass were pooled, lyophilized, and stored at –80°C.

To assemble full-length NFL (residues 2 to 543), native chemical ligation reactions were performed by combining peptides bearing C-terminal SEA moieties (NFL 2 to 26, WT, P8L, P8esL, P8meL, P8Q, P8meQ, P8R, P8meR, P8dmP, P22S, P22meS, P22R, P22meR, P22T, and P22meT) with the NFL truncated fragment bearing N-terminal cysteine (S27C, residues 27 to 543). Ligation reactions were performed as described for isolated NFL head domain assembly.

Assembly of semisynthetic hnRNPA2 LCD constructs

To assemble the hnRNPA2 LCD (residues 181 to 341) a three-piece native chemical ligation strategy was devised. In ligation 1, peptides bearing an N-terminal Thz and C-terminal SEA moiety (hnRNPA2 residues 285 to 305, hnRNPA2 residues 285 to 305 with P298L mutation, or hnRNPA2 residues 285 to 305 with P298meL mutation) were combined with the hnRNPA2 C-terminal peptide (residues 306 to 341, S306C). Reactions included 1 mM SEA peptide, 0.5 mM C-amidated peptide, 200 mM MES-Na, 20 mM TCEP, and 150 mM TFET in a degassed buffer of 6 M guanidine-HCl and 0.1 M sodium phosphate at pH 7.0. Reactions were incubated at 37°C for 16 hours, and progress was monitored by RP-HPLC and ESI-MS analysis. Upon completion of the reaction, Thz deprotection was initiated by the addition of 200 mM O-methylhydroxylamine-HCl and 20 mM TCEP in a degassed buffer of 6 M guanidine-HCl and 0.1 M sodium phosphate. Reactions were adjusted to pH 4.0 and incubated at 37°C for 4 hours. Products were purified on a preparative C18 RP-HPLC column and analyzed by analytical RP-HPLC and ESI-MS. Fractions containing target mass were pooled, lyophilized, and stored at –80°C. These products, WT, P298L, and P298meL, are referred to as hnRNPA2 piece 2+3.

For ligation 2, the hnRNPA2 piece 2+3 constructs, each bearing a free N-terminal cysteine, were combined with the recombinant hnRNPA2 thioester construct (residues 181 to 284). These reactions included 1 mM piece 2+3, 0.5 mM recombinant hnRNPA2 thioester fragment, 20 mM TCEP, and 150 mM TFET in a degassed buffer of 6 M guanidine-HCl and 0.1 M sodium phosphate at pH 7.0. Reactions were incubated at 37°C for 16 hours, and progress was monitored using RP-HPLC and ESI-MS analysis. Upon completion of the reactions, the ligation products were purified on a semipreparative C18 RP-HPLC column and fractions were characterized by RP-HPLC and ESI-MS. Pure fractions were pooled, lyophilized, and stored at –80°C.

TDP-43 LCD phase-separated droplet formation

All His-TDP-43 LCD WT constructs, mutation variants, and single N^{α}-methyl variants were dissolved in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 6 M guanidine-HCl, 10 mM β-mercaptoethanol (β-ME) and diluted to 300 μM with same buffer. Because the yield of triple N^{α}-methyl variants is low, triple N^{α}-methyl variants (triple-me (321-323-325), triple-me (323-325-327)) and WT His-TDP-43 LCD were dissolved and diluted to 100 μM with the same buffer. For glycine scanning mutagenesis experiments, droplet formation was induced by diluting all constructs 30-fold in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 10 mM β-ME in the presence of 0, 0.4, 0.8, or 1.2 M urea. Solutions were loaded onto a clear-bottomed Corning Costa 384-well plate and imaged on a ZOE Fluorescent Cell Imager (Bio-Rad).

For N^{α}-methyl scanning experiments, droplet formation was induced by diluting all constructs 30-fold in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM β-ME, and 5 μM $NiCl_2$ in the presence of 0, 0.4, 0.8, or 1.2 M urea. WT and triple N^{α}-methyl variants were diluted 10-fold to a 10 μM final concentration with the droplet formation buffer used in N^{α}-methyl scanning experiments. The formed droplets were immediately loaded onto a clear-bottomed 384-well plate, and absorbance at 600 nm was measured to determine turbidity. Triplicate measurements were performed for all constructs at each condition. After turbidity measurements, droplets were also imaged on a ZOE Fluorescent Cell Imager.

Transient transfection of U2OS cells

U2OS cells were cultured in a Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). The U2OS cells were seeded on a 35-mm glass-bottomed confocal dish 1 day before transfection. Cells were transfected with FLAG-GFP-TDP-43 plasmid (or mutations thereof) using Lipofectamine 3000 (Thermo Fisher). After 24 hours, cells were washed twice with phosphate-buffered saline (PBS), incubated with Hoechst staining reagent (diluted in PBS) for 10 min, and imaged by confocal microscopy. Cells were also harvested in SDS sample loading buffer for Western blotting (α-FLAG-HRP, Millipore-Sigma) to confirm similar expression levels of each TDP-43 variant.

Creation of TDP-43 inducible expression cell line and cell viability assay

All constructs were transfected into a Tet-inducible cell line, U2OS-TR. The U2OS-TR cells were grown in Tet-negative FBS to minimize leaky expression. Cells (1×10^6) were plated 24 hours before transfection. DNA (1 μg) was transfected with Lipofectamine 3000 (Thermo Fisher) according to the manufacturer's protocol. After 24 hours, cells were split into 10 × 100 mm plates in medium containing G418 at a concentration of 700 μg/ml. Selection continued for ~3 weeks (refreshing G418 twice per week). Single colonies were picked using clonal rings, propagated, and doxycycline-induced expression levels were analyzed. For FLAG-GFP-TDP-43 induction, doxycycline was added to a concentration of 1 μg/ml for 24 hours. Clonal populations were harvested in SDS sample loading buffer for Western blotting (α-FLAG) to confirm similar expression levels of each TDP-43 variant. Colonies from each variant that exhibited similar FLAG-GFP-TDP-43 expression levels were used in the viability assays.

Each stable expression cell line was plated in quadruplet into 7 × 96-well plates at 4000 cells/well. Induction proceeded as previously mentioned using doxycycline starting 24 hours after plating. Plates were assayed for cell viability using CellTiter-Glo (Promega) according to the manufacturer's protocol (CellTiter-Glo/Promega) each day for 7 days.

Phenotypic analysis of NFL filament assembly, NFL N-terminal peptide solubility, and head domain polymer formation

NFL filament assembly was performed as previously described (*43*). To test NFL peptide solubility, the lyophilized peptides were first completely dissolved with neat TFA, which was then evaporated under a gentle stream of N_2. Residual TFA was removed by lyophilization for 30 min. This step is important to ensure that peptides are completely disaggregated before ThT analysis (*68*). Lyophilized peptides were then dissolved in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM β-ME, and 6 M guanidine-HCl, and then diluted into a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM β-ME, and 20 μM ThT (final peptide concentration = 200 μM). Solutions were next loaded into a 384-well fluorescence assay plate (50 μl per well), sealed, and ThT fluorescence was monitored for 99 hours (Cytation 5 plate reader, BioTek). Five repeats were performed for all samples.

To test NFL head domain polymer formation, each variant was dissolved in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM β-ME, and 8 M urea, and then dialyzed against an identical buffer with a reduced urea concentration (3 M urea). Sample turbidity was measured by absorbance at 600 nm, and head domain polymers were confirmed by negative-staining EM.

ThT assay for tau peptides

Lyophilized tau peptides were first completely dissolved with neat TFA, which was then evaporated under a gentle stream of N_2. Residual TFA was removed by lyophilization for 30 min (*52*). Next, peptides were dissolved in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 6 M guanidine-HCl, and then diluted into PBS supplemented with 20 μM ThT (final peptide concentration = 100 μM). Solutions were loaded into a 96-well fluorescence assay plate, sealed (Constar 6570 sealing tape), and ThT fluorescence was monitored for 99 hours (Cytation 5 plate reader, BioTek). Five repeats were performed for all samples.

Phenotypic analysis of tau peptides in tau biosensor cell line

The tau biosensor cell line was purchased from ATCC and cultured in DMEM supplemented with 10% FBS. Tau biosensor cells were seeded in a 96-well tissue culture plate the day before tau peptide transfection.

Tau peptides were disaggregated, diluted in PBS to 300 μM, and sonicated in a water bath sonicator (Branson 2800) for 2 min. Tau peptide solution was mixed with same volume of a lipofectamine 3000-OptiMEM solution and incubated for 30 min at room temperature. Ten microliters of tau peptide transfection mixture was added to 90 μl of tau biosensor cells to reach a final peptide concentration of 15 μM in medium. The cells were imaged with an EVOS FL fluorescence microscope 48 to 72 hours after transfection.

Phenotypic analysis of hnRNPA2 LCD phase-separated droplets

The hnRNPA2 LCD proteins were dissolved in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 6 M guanidine-HCl. Then, hnRNPA2 LCD phase-separated droplet formation was induced by diluting protein in a buffer containing either 25 mM Tris (pH 7.5), 150 mM NaCl, and 5 mM β-ME or 50 mM MES (pH 5.5), 100 mM NaCl, and 5 mM β-ME to a final concentration of 15 μM. Identical results were observed for the WT hnRNPA2 LCD and its variants in these two buffers. However, the average size of hnRNPA2 droplets increased in the pH 5.5 buffer, and these conditions were used for image analysis with the ZOE Fluorescent Cell Imager.

ThT assay for hnRNPA2 LCD peptides

Synthesized hnRNPA2 LCD peptides were dissolved in a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 6 M guanidine-HCl. The dissolved peptides were diluted into a buffer containing 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 20 μM ThT (final peptide concentration = 400 μM), loaded in a 384-well fluorescence assay plate (50 μl per well), and sealed. ThT fluorescence was monitored for 99 hours using the Cytation 5 plate reader (BioTek). Five repeats were performed for all samples.

REFERENCES AND NOTES

1. M. E. Oates *et al.*, D^2P^2: database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508–D516 (2013). doi: 10.1093/nar/gks1226; pmid: 23203878
2. J. Ma, M. Ptashne, Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48**, 847–853 (1987). doi: 10.1016/0092-8674(87)90081-X; pmid: 3028647
3. S. J. Triezenberg, R. C. Kingsbury, S. L. McKnight, Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* **2**, 718–729 (1988). doi: 10.1101/gad.2.6.718; pmid: 2843425
4. A. J. Courey, D. A. Holtzman, S. P. Jackson, R. Tjian, Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59**, 827–836 (1989). doi: 10.1016/0092-8674(89)90606-5; pmid: 2512012
5. V. Pejaver *et al.*, The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci.* **23**, 1077–1093 (2014). doi: 10.1002/pro.2494; pmid: 24888500
6. J. Woodsmith, A. Kamburov, U. Stelzl, Dual coordination of post translational modifications in human protein networks. *PLOS Comput. Biol.* **9**, e1002933 (2013). doi: 10.1371/journal.pcbi.1002933; pmid: 23505349
7. S. Frey, D. Görlich, A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* **130**, 512–523 (2007). doi: 10.1016/j.cell.2007.06.024; pmid: 17693259
8. M. Kato *et al.*, Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012). doi: 10.1016/j.cell.2012.04.017; pmid: 22579281
9. S. Xiang *et al.*, The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* **163**, 829–839 (2015). doi: 10.1016/j.cell.2015.10.040; pmid: 26544936
10. A. Molliex *et al.*, Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015). doi: 10.1016/j.cell.2015.09.015; pmid: 26406374
11. A. Patel *et al.*, A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015). doi: 10.1016/j.cell.2015.07.047; pmid: 26317470
12. S. Das, M. Vera, V. Gandin, R. H. Singer, E. Tutucci, Intracellular mRNA transport and localized translation. *Nat. Rev. Mol. Cell Biol.* **22**, 483–504 (2021). doi: 10.1038/s41580-021-00356-8; pmid: 33837370
13. N. H. Alami *et al.*, Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. *Neuron*

81, 536–543 (2014). doi: 10.1016/j.neuron.2013.12.018; pmid: 24507191

14. J. F. Chu, P. Majumder, B. Chatterjee, S. L. Huang, C. J. Shen, TDP-43 regulates coupled dendritic mRNA transport-translation processes in co-operation with FMRP and Staufen1. *Cell Rep.* **29**, 3118–3133.e6 (2019). doi: 10.1016/j.celrep.2019.10.061; pmid: 31801077
15. R. K. Narayanan *et al*., Identification of RNA bound to the TDP-43 ribonucleoprotein complex in the adult mouse brain. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **14**, 252–260 (2013). doi: 10.3109/21678421.2012.734520; pmid: 23134510
16. T. Arai *et al*., TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006). doi: 10.1016/j.bbrc.2006.10.093; pmid: 17084815
17. M. Neumann *et al*., Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006). doi: 10.1126/science.1134108; pmid: 17023659
18. L. Lim, Y. Wei, Y. Lu, J. Song, ALS-causing mutations significantly perturb the self-assembly and interaction with nucleic acid of the intrinsically disordered prion-like domain of TDP-43. *PLOS Biol.* **14**, e1002338 (2016). doi: 10.1371/journal.pbio.1002338; pmid: 26735904
19. Y. Lin *et al*., Redox-mediated regulation of an evolutionarily conserved cross-β structure formed by the TDP43 low complexity domain. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28727–28734 (2020). doi: 10.1073/pnas.2012216117; pmid: 33144500
20. B. Portz, B. L. Lee, J. Shorter, FUS and TDP-43 phases in health and disease. *Trends Biochem. Sci.* **46**, 550–563 (2021). doi: 10.1016/j.tibs.2020.12.005; pmid: 33446423
21. T. Murakami *et al*., ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* **88**, 678–690 (2015). doi: 10.1016/j.neuron.2015.10.030; pmid: 26526393
22. B. S. Johnson *et al*., TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J. Biol. Chem.* **284**, 20329–20339 (2009). doi: 10.1074/jbc.M109.010264; pmid: 19465477
23. E. Buratti, Functional significance of TDP-43 mutations in disease. *Adv. Genet.* **91**, 1–53 (2015). doi: 10.1016/bs.adgen.2015.07.001; pmid: 26410029
24. H. B. Schmidt, D. Görlich, Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *eLife* **4**, e04251 (2015). doi: 10.7554/eLife.04251; pmid: 25562883
25. Y. Lin, S. L. Currie, M. K. Rosen, Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* **292**, 19110–19120 (2017). doi: 10.1074/jbc.M117.800466; pmid: 28924037
26. E. W. Martin, T. Mittag, Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry* **57**, 2478–2487 (2018). doi: 10.1021/acs.biochem.8b00008; pmid: 29517898
27. Q. Cao, D. R. Boyer, M. R. Sawaya, P. Ge, D. S. Eisenberg, Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat. Struct. Mol. Biol.* **26**, 619–627 (2019). doi: 10.1038/s41594-019-0248-4; pmid: 31235914
28. E. L. Guenther *et al*., Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat. Struct. Mol. Biol.* **25**, 463–471 (2018). doi: 10.1038/s41594-018-0064-2; pmid: 29786080
29. D. T. Murray *et al*., Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **171**, 615–627.e16 (2017). doi: 10.1016/j.cell.2017.08.048; pmid: 28942918
30. L. L. Jiang *et al*., Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J. Biol. Chem.* **288**, 19614–19624 (2013). doi: 10.1074/jbc.M113.463828; pmid: 23689371
31. A. E. Conicella, G. H. Zerze, J. Mittal, N. L. Fawzi, ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **24**, 1537–1549 (2016). doi: 10.1016/j.str.2016.07.007; pmid: 27545621
32. Q. Li, W. M. Babinchak, W. K. Surewicz, Cryo-EM structure of amyloid fibrils formed by the entire low complexity domain of TDP-43. *Nat. Commun.* **12**, 1620 (2021). doi: 10.1038/s41467-021-21912-y; pmid: 33712624
33. A. Jordanova *et al*., Mutations in the neurofilament light chain gene (NEFL) cause early onset severe Charcot-Marie-Tooth disease. *Brain* **126**, 590–597 (2003). doi: 10.1093/brain/awg059; pmid: 12566280
34. J. S. Shin *et al*., NEFL Pro22Arg mutation in Charcot-Marie-Tooth disease type 1. *J. Hum. Genet.* **53**, 936–940 (2008). doi: 10.1007/s10038-008-0333-8; pmid: 18758688
35. R. Perez-Olle, S. T. Jones, R. K. Liem, Phenotypic analysis of neurofilament light gene mutations linked to Charcot-Marie-Tooth disease in cell culture models. *Hum. Mol. Genet.* **13**, 2207–2220 (2004). doi: 10.1093/hmg/ddh236; pmid: 15282209
36. R. Pérez-Ollé *et al*., Mutations in the neurofilament light gene linked to Charcot-Marie-Tooth disease cause defects in transport. *J. Neurochem.* **93**, 861–874 (2005). doi: 10.1111/j.1471-4159.2005.03095.x; pmid: 15857389
37. G. Y. Ching, R. K. Liem, Analysis of the roles of the head domains of type IV rat neuronal intermediate filament proteins in filament assembly using domain-swapped chimeric proteins. *J. Cell Sci.* **112**, 2233–2240 (1999). doi: 10.1242/jcs.112.13.2233; pmid: 10362553
38. S. R. Gill, P. C. Wong, M. J. Monteiro, D. W. Cleveland, Assembly properties of dominant and recessive mutations in the small mouse neurofilament (NF-L) subunit. *J. Cell Biol.* **111**, 2005–2019 (1990). doi: 10.1083/jcb.111.5.2005; pmid: 2121744
39. A. Petzold, Neurofilament phosphoforms: Surrogate markers for axonal injury, degeneration and loss. *J. Neurol. Sci.* **233**, 183–198 (2005). doi: 10.1016/j.jns.2005.03.015; pmid: 15896809
40. M. Kornreich, R. Avinery, E. Malka-Gibor, A. Laser-Azogui, R. Beck, Order and disorder in intermediate filament proteins. *FEBS Lett.* **589** (19PartA), 2464–2476 (2015). doi: 10.1016/j.febslet.2015.07.024; pmid: 26231765
41. H. Herrmann, U. Aebi, Intermediate filaments: Structure and assembly. *Cold Spring Harb. Perspect. Biol.* **8**, a018242 (2016). doi: 10.1101/cshperspect.a018242; pmid: 27803112
42. Y. Lin *et al*., Toxic PR poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers. *Cell* **167**, 789–802.e12 (2016). doi: 10.1016/j.cell.2016.10.003; pmid: 27768897
43. X. Zhou *et al*., Transiently structured head domains control intermediate filament assembly. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022121118 (2021). doi: 10.1073/pnas.2022121118; pmid: 33593918
44. M. Kato *et al*., Redox state controls phase separation of the yeast ataxin-2 Protein via reversible oxidation of its methionine-rich low-complexity domain. *Cell* **177**, 711–721.e8 (2019). doi: 10.1016/j.cell.2019.02.044; pmid: 30982603
45. S. S. A. An *et al*., Retention of the cis proline conformation in tripeptide fragments of bovine pancreatic ribonuclease A containing a non-natural proline analogue, 5,5-dimethylproline. *J. Am. Chem. Soc.* **121**, 11558–11566 (1999). doi: 10.1021/ja9930317
46. O. Bugiani *et al*., Frontotemporal dementia and corticobasal degeneration in a family with a P301S mutation in tau. *J. Neuropathol. Exp. Neurol.* **58**, 667–677 (1999). doi: 10.1097/00005072-199906000-00011; pmid: 10374757
47. L. N. Clark *et al*., Pathogenic implications of mutations in the tau gene in pallido-ponto-nigral degeneration and related neurodegenerative disorders linked to chromosome 17. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13103–13107 (1998). doi: 10.1073/pnas.95.22.13103; pmid: 9789048
48. A. D. Sperfeld *et al*., FTDP-17: An early-onset phenotype with parkinsonism and epileptic seizures caused by a novel mutation. *Ann. Neurol.* **46**, 708–715 (1999). doi: 10.1002/1531-8249(199911)46:5<708::AID-ANA5>3.0.CO;2-K; pmid: 10553987
49. A. Lossos *et al*., Frontotemporal dementia and parkinsonism with the P301S tau gene mutation in a Jewish family. *J. Neurol.* **250**, 733–740 (2003). doi: 10.1007/s00415-003-1074-4; pmid: 12796837
50. A. Lladó *et al*., A novel MAPT mutation (P301T) associated with familial frontotemporal dementia. *Eur. J. Neurol.* **14**, e9–e10 (2007). doi: 10.1111/j.1468-1331.2007.01763.x; pmid: 17662000
51. K. H. Strang *et al*., Distinct differences in prion-like seeding and aggregation between Tau protein variants provide mechanistic insights into tauopathies. *J. Biol. Chem.* **293**, 4579 (2018). doi: 10.1074/jbc.AAC118.002657; pmid: 29572329
52. D. Chen *et al*., Tau local structure shields an amyloid-forming motif and controls aggregation propensity. *Nat. Commun.* **10**, 2493 (2019). doi: 10.1038/s41467-019-10355-1; pmid: 31175300
53. C. Renner *et al*., Fluoroprolines as tools for protein design and engineering. *Angew. Chem. Int. Ed.* **40**, 923–925 (2001). doi: 10.1002/1521-3773(20010302)40:5<923::AID-ANIE923>3.0.CO;2-#
54. B. B. Holmes *et al*., Proteopathic tau seeding predicts tauopathy in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4376–E4385 (2014). doi: 10.1073/pnas.1411649111; pmid: 25261551
55. P. M. Seidler *et al*., Structure-based inhibitors halt prion-like seeding by Alzheimer's disease-and tauopathy-derived brain tissue samples. *J. Biol. Chem.* **294**, 16451–16464 (2019). doi: 10.1074/jbc.RA119.009688; pmid: 31537646
56. X. Qi *et al*., Familial early-onset Paget's disease of bone associated with a novel hnRNPA2B1 mutation. *Calcif. Tissue Int.* **101**, 159–169 (2017). doi: 10.1007/s00223-017-0269-0; pmid: 28389692
57. D. T. Murray *et al*., Structural characterization of the D290V mutation site in hnRNPA2 low-complexity-domain polymers. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9782–E9791 (2018). doi: 10.1073/pnas.1806174115; pmid: 30279180
58. J. Lu *et al*., CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. *Nat. Commun.* **11**, 4090 (2020). doi: 10.1038/s41467-020-17905-y; pmid: 32796831
59. V. H. Ryan *et al*., Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol. Cell* **69**, 465–479.e7 (2018). doi: 10.1016/j.molcel.2017.12.022; pmid: 29358076
60. Y. S. Yang *et al*., Yeast ataxin-2 forms an intracellular condensate required for the inhibition of TORC1 signaling during respiratory growth. *Cell* **177**, 697–710.e17 (2019). doi: 10.1016/j.cell.2019.02.043; pmid: 30982600
61. V. O. Sysoev *et al*., Dynamic structural order of a low-complexity domain facilitates assembly of intermediate filaments. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23510–23518 (2020). doi: 10.1073/pnas.2010000117; pmid: 32907935
62. J. Chatterjee, F. Rechenmacher, H. Kessler, N-methylation of peptides and proteins: An important element for modulating biological functions. *Angew. Chem. Int. Ed.* **52**, 254–269 (2013). doi: 10.1002/anie.201205674; pmid: 23161799
63. E. W. Martin *et al*., Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020). doi: 10.1126/science.aaw8653; pmid: 32029630
64. M. P. Hughes *et al*., Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698–701 (2018). doi: 10.1126/science.aan6398; pmid: 29439243
65. M. Kato, S. L. McKnight, The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2114412118 (2021). doi: 10.1073/pnas.2114412118; pmid: 34654750
66. X. Gui *et al*., Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* **10**, 2006 (2019). doi: 10.1038/s41467-019-09902-7; pmid: 31043593
67. N. Ollivier *et al*., Tidbits for the synthesis of bis(2-sulfanylethyl) amido (SEA) polystyrene resin, SEA peptides and peptide thioesters. *J. Pept. Sci.* **20**, 92–97 (2014). doi: 10.1002/psc.2580; pmid: 24254655
68. B. O'Nuallain *et al*., Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay. *Methods Enzymol.* **413**, 34–74 (2006). doi: 10.1016/S0076-6879(06)13003-7; pmid: 17046390

ACKNOWLEDGMENTS

We thank D. Nijhawan and B. Tu for scientific input and encouragement; R. Thompson for technical advice; J. Mohapatra for assistance with peptide synthesis; and U. Shibler, R. Losick, B. Alberts, M. Brown, and J. Goldstein for assistance in producing the manuscript. This work was supported by an anonymous donor (funds to S.L.M.); the National Institute of General Medical Science, National Institutes of Health (NIH) (grant GM130358 to S.L.M.); the National Cancer Institute, NIH (grant CA231649 to S.L.M.); the Welch Foundation (grant I-2039-20200401 to G.L. and

grant I-2020-20190330 to T.Q.); and the Cancer Prevention Research Institute of Texas (grant RR180051 to G.L.). **Author contributions:** Conceptualization: S.L.M., G.L., and X.Z. Investigation: X.Z., L.S., K.T., L.S. and D.L. Methodology: G.L., S.L.M., T.Q., X.Z., and K.T. Visualization: X.Z., G.L., S.L.M. and M.K. Writing – original draft: S.L.M. and G.L. Writing – review and editing: S.L.M., G.L., X.Z. and K.T. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn5582
Materials and Methods
Figs. S1 to S6
References
Data S1 and S2
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 4 December 2021; accepted 6 May 2022
10.1126/science.abn5582

RESEARCH ARTICLE

NUTRIENT METABOLISM

Zonated leucine sensing by Sestrin-mTORC1 in the liver controls the response to dietary leucine

Andrew L. Cangelosi[1,2,3]*, Anna M. Puszynska[1,2], Justin M. Roberts[1,2,3], Andrea Armani[1,2,4,5], Thao P. Nguyen[1,2,3], Jessica B. Spinelli[1,2], Tenzin Kunchok[1], Brianna Wang[1], Sze Ham Chan[1]†, Caroline A. Lewis[1], William C. Comb[1,2]‡, George W. Bell[1], Aharon Helman[6], David M. Sabatini[3]§

The mechanistic target of rapamycin complex 1 (mTORC1) kinase controls growth in response to nutrients, including the amino acid leucine. In cultured cells, mTORC1 senses leucine through the leucine-binding Sestrin proteins, but the physiological functions and distribution of Sestrin-mediated leucine sensing in mammals are unknown. We find that mice lacking Sestrin1 and Sestrin2 cannot inhibit mTORC1 upon dietary leucine deprivation and suffer a rapid loss of white adipose tissue (WAT) and muscle. The WAT loss is driven by aberrant mTORC1 activity and fibroblast growth factor 21 (FGF21) production in the liver. Sestrin expression in the liver lobule is zonated, accounting for zone-specific regulation of mTORC1 activity and FGF21 induction by leucine. These results establish the mammalian Sestrins as physiological leucine sensors and reveal a spatial organization to nutrient sensing by the mTORC1 pathway.

Leucine is an essential amino acid needed to synthesize proteins and metabolites such as branched chain fatty acids (*1–3*). In addition, it has been recognized for decades that leucine has distinctive physiological effects, including promoting skeletal muscle growth (*4–8*), insulin secretion (*9–11*), and immune function (*12–14*) and modulating health span and life span in mice (*15–17*). Moreover, plasma leucine concentrations are also implicated in certain pathological states, such as metabolic syndrome (*18–20*).

A key effector of leucine is thought to be the mechanistic target of rapamycin complex 1 (mTORC1) protein kinase, a master regulator of growth and metabolism. Diverse nutrients, growth factors, and stresses regulate mTORC1 (*4*, *21–23*), and understanding how it detects so many inputs has been of longstanding interest. A model has started to emerge of the nutrient-sensing mechanisms upstream of mTORC1 in which nutrient-derived signals converge on the heterodimeric Rag guanosine triphosphatases (GTPases) and their many regulators, including the GATOR1, GATOR2, and Ragulator complexes. The Rag heterodimer binds mTORC1 in a nutrient-sensitive manner to control its localization to the lysosomal surface, where it can interact with its kinase activator, the Rheb GTPase (*4*, *21–23*).

How the mTORC1 pathway senses leucine has been highly debated (*24–31*). Several years ago, we showed that in cultured cells the leucine-binding proteins Sestrin1 and Sestrin2 serve as leucine sensors for the pathway (*32*, *33*). Growing evidence implicates the Sestrins in various facets of organismal function (*34–39*). However, whether the mammalian Sestrins have a leucine-sensing role in vivo—and, if so, in which tissues they act and the physiology they control as leucine sensors—is unknown.

Results

Sestrin1 and Sestrin2 are physiological leucine sensors for mTORC1

To study leucine sensing by mTORC1 in vivo and minimize confounding effects of other nutrient alterations, we tested the response of mice to changes in dietary leucine content. We first fasted and refed mice with food containing 100, 10, or 0% of the leucine content of standard chow (Fig. 1A). Feeding of these diets caused stepwise reductions in plasma leucine concentrations (Fig. 1B) and in the phosphorylation of the mTORC1 substrates S6K1 (S6 kinase 1) and 4EBP1 (eukaryotic translation initiation factor 4E–binding protein 1) in the liver (Fig. 1C), showing that, in our experimental system, dietary leucine controls mTORC1 activity in vivo.

In cultured cells, Sestrin1 and Sestrin2 inhibit mTORC1 signaling by interacting with and suppressing—in a leucine-sensitive manner—the GATOR2 complex, a positive component of the mTORC1 pathway (*32*). To determine whether the same interaction occurs in vivo, we immunoprecipitated, from the liver of mice, GATOR2 using an antibody to its WDR24 component. Consistent with leucine disrupting the Sestrin1/2-GATOR2 interaction, GATOR2 coimmunoprecipitated greater amounts of Sestrin1 and Sestrin2 in mice refed the leucine-free than in those fed the control diet (Fig. 1D). Notably, the addition of leucine, but not arginine, to the immunopurified complexes disrupted the Sestrin1/2-GATOR2 interaction (Fig. 1D). Thus, as in cultured cells, leucine regulates the binding of Sestrin1 and Sestrin2 to GATOR2 in vivo in mouse tissues.

To determine whether the regulation of mTORC1 by dietary leucine requires Sestrin1 and Sestrin2, we generated mice lacking both proteins [double-knockout (DKO) mice] and fasted and refed them with the control or leucine-free diet. Whereas mTORC1 activity in the liver was low in wild-type (WT) mice refed with the leucine-free diet, in DKO mice, mTORC1 activity was high, irrespective of the leucine content of the diet in both males (Fig. 1E) and females (fig. S1A). Sestrin expression did not affect food consumption (fig. S1B). Loss of either Sestrin1 or Sestrin2 alone had no impact on the sensitivity of mTORC1 to leucine deprivation (fig. S1, C and D), consistent with Sestrin1 and Sestrin2 having redundant functions in cultured cells (*32*). As it does in the liver, leucine regulates mTORC1 activity in white adipose tissue (WAT) in a Sestrin-dependent manner in males (Fig. 1F) and females (fig. S1E).

mTORC1 signaling remained sensitive to fasting in the liver and WAT of DKO mice (fig. S2, A and B) and to starvation of all amino acids (but not to starvation of only leucine) in primary hepatocytes and WAT explants obtained from DKO mice (fig. S2, C and D). Thus, loss of the Sestrins affects the response of mTORC1 specifically to leucine deprivation.

The Sestrins have been proposed to impinge on the mTORC1 pathway through regulation of adenosine monophosphate–activated protein kinase (AMPK) (*36*). However, in our model system, liver-specific deletion of both AMPK catalytic subunits (AMPKα1 and AMPKα2) did not affect the activation of hepatic mTORC1 by dietary leucine, despite eliminating AMPK activity, as indicated by the absence of phosphorylation of acetyl–coenzyme A carboxylase (ACC), a canonical AMPK substrate (fig. S2E).

On the basis of the structure of the leucine-binding pocket of Sestrin2, we previously identified a point mutation [Trp[444]→Leu (W444L)] that reduces, but does not abolish, the affinity

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [2]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Veneto Institute of Molecular Medicine, 35129 Padova, Italy. [5]Department of Biomedical Sciences, University of Padova, 35131 Padova, Italy. [6]Institute of Biochemistry, Food Science and Nutrition, Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot 7610001, Israel.
***Corresponding author. Email: andrewcangelosi1@gmail.com**
†Present address: Department of Pharmacology, University of Virginia, Charlottesville, VA 22903, USA. ‡Present address: Mythic Therapeutics, Waltham, MA 02453, USA. §David M. Sabatini is no longer affiliated with the Whitehead Institute, the Howard Hughes Medical Institute, or Massachusetts Institute of Technology. To ensure execution of the duties of corresponding author, Andrew Cangelosi has taken on this role.

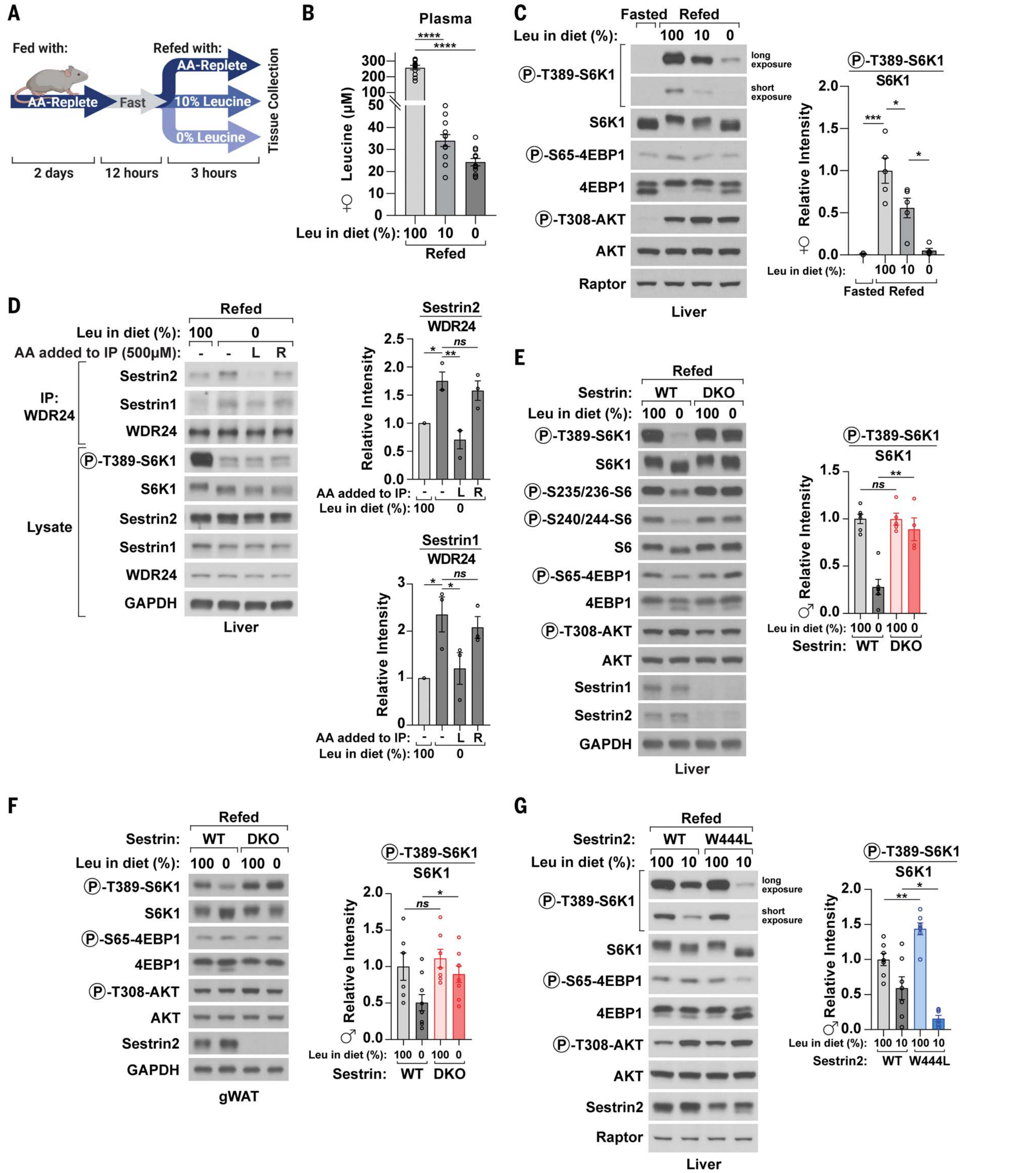


Fig. 1. The Sestrins control leucine sensing by mTORC1 in vivo.
(**A**) Schematic of the experimental setup for studying leucine sensing in vivo. Mice maintained on an amino acid (AA)–replete control diet for 2 days were fasted overnight for 12 hours then refed with food containing the indicated leucine contents, and tissues were collected 3 hours after the start of the feeding period. (**B**) Plasma leucine concentrations in wild-type female mice 3 hours after eating the indicated diets (n = 11 to 12 mice). (**C**) Phosphorylation state and amounts of indicated proteins in liver lysates from wild-type female mice refed with the indicated diets for 3 hours (n = 3 to 5 mice). (**D**) Dietary leucine regulates Sestrin-GATOR2 interactions. Endogenous WDR24 immunoprecipitates (IPs) were prepared from liver lysates from wild-type male mice refed with the indicated diets for 3 hours. In lanes 2 to 4, IPs were prepared from equal volumes of the same liver lysate, and, where noted, indicated amino acids were added during washes. L, leucine; R, arginine; GAPDH, glyceraldehyde phosphate dehydrogenase. IPs and liver lysates were analyzed by immunoblotting for the phosphorylation states and amounts of the indicated proteins (n = 3 mice). (**E**) Male mice with indicated genotypes were refed with the indicated diets for 3 hours. Liver lysates were analyzed by immunoblotting for the phosphorylation state and amounts of the indicated proteins (n = 4 to 6 mice). (**F**) Gonadal WAT (gWAT) lysates from

male mice treated as in (E) were analyzed by immunoblotting for the phosphorylation states and amounts of the indicated proteins (n = 6 to 9 mice). (**G**) Female mice with the indicated liver genotypes were refed with diets with different leucine contents for 3 hours, and liver lysates were analyzed by immunoblotting for the phosphorylation state and amounts of the indicated proteins (n = 7 mice). Data are the mean ± SEM. P values were determined using two-tailed t tests [(B) and (E) to (G)], one-way analysis of variance (ANOVA) with Tukey test [(B) and (C)], or one-way ANOVA with Dunnett's test (D). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. ns, not significant.

of Sestrin2 for leucine (*33*). To determine how this mutation affects the activation of mTORC1 by leucine in vivo, we generated knockin mice expressing Sestrin2 W444L from the endogenous *Sesn2* locus (*Sesn2*^{W444L} mice). Leucine, over a range of concentrations, activated mTORC1 to a lesser degree in primary hepatocytes from *Sesn2*^{W444L} than control (*Sesn2*^{WT}) mice (fig. S2F). Similarly, food containing 10% of the leucine content of normal chow activated mTORC1 to a lesser extent in the livers of *Sesn2*^{W444L} than *Sesn2*^{WT} mice (Fig. 1G). Thus, the affinity of Sestrin2 for leucine determines the sensitivity of mTORC1 to the leucine content of the diet. Taken together, our results establish that, as in cultured cells, Sestrin1 and Sestrin2 transmit leucine availability to the mTORC1 pathway in vivo.

Sestrin-mediated leucine sensing preserves WAT and muscle mass during dietary leucine deprivation

To examine the physiological importance of leucine sensing by mTORC1, we fed wild-type and DKO mice lacking Sestrin1 and Sestrin2 the leucine-free diet for 8 days (Fig. 2A). DKO mice, but not *Sesn1*$^{-/-}$ mice or *Sesn2*$^{-/-}$ mice, lost more body weight than wild-type controls (Fig. 2B and fig. S3) despite eating similar amounts of the leucine-free food (fig. S4). The greater reduction in body weight in DKO mice was a consequence of a severe loss of WAT, as evident in gross and microscopic examinations of several fat depots, as well as a reduction in skeletal muscle mass (Fig. 2, C to G). We saw these phenotypes in both sexes (fig. S5), although the WAT loss was more pronounced in female mice. Liver mass did not contribute to the difference in body weight, as it was unaffected by Sestrin loss (fig. S6). Although wild-type and DKO mice ate a similarly reduced amount of the leucine-free food in comparison with control food (fig. S4), this reduction in food intake does not account for the different responses of wild-type and DKO mice to leucine-free feeding (fig. S7, A to C). Furthermore, the reductions in body weight, WAT, and skeletal muscle occurred in leucine-deprived DKO mice regardless of whether they were fasted before feeding with the leucine-free diet (fig. S7, D to F). As mTORC1 activity and leucine deprivation can both affect insulin signaling and glucose homeostasis (*4*, *40*, *41*), we assessed glucose tolerance, insulin secretion, and hepatic insulin sensitivity but found no alterations in the DKO mice (fig. S8). On the leucine-free diet, wild-type and DKO mice had similarly low levels of plasma leucine (Fig. 2H and fig. S9). However, whereas plasma levels of several other amino acids were either maintained or in some cases even increased in wild-type mice, levels of those same amino acids were reduced in DKO mice (Fig. 2H and fig. S9).

Treatment with rapamycin, an mTORC1 inhibitor, restored the total body weight and WAT mass of DKO mice to those of wild-type animals (fig. S10). Aberrant mTORC1 activity thus underlies the inappropriate response of DKO mice to leucine deprivation.

Notably, on diets lacking valine or methionine, DKO and wild-type mice lost equal amounts of total body weight, WAT, and skeletal muscle (Fig. 2, I to N). The lack of altered responses in DKO mice to these diets is consistent with the Sestrins being specific sensors of leucine at physiological amino acid concentrations (*32*).

Lastly, *Sesn2*^{W444L} mice, in which mTORC1 is inhibited more strongly upon leucine deprivation than in wild-type animals (Fig. 1G), lost less WAT mass than control *Sesn2*^{WT} mice over an extended period (16 days) of leucine deprivation (Fig. 2, O to Q), despite having similar intake of the leucine-free food (fig. S11A). Thus, the leucine-binding capacity of Sestrin2 modulates the physiological response to leucine deprivation. In contrast, leucine-deprived *Sesn2*^{W444L} and *Sesn2*^{WT} mice lost similar amounts of skeletal muscle (fig. S11, B to D), consistent with muscle not expressing detectable levels of Sestrin2 (fig. S11E) (*42*). We conclude that mice require the Sestrin-mediated regulation of mTORC1 to maintain homeostasis specifically in response to limited leucine availability.

Leucine sensing in the liver controls the response of WAT to dietary leucine deprivation through FGF21

Because adipocyte-specific hyperactivation of mTORC1 can lead to a reduction in WAT mass (*43*), we considered the possibility that mTORC1 deregulation in the adipocytes of the WAT itself (Fig. 1F and figs. S1E and S2D) might cause its loss in DKO mice on the leucine-free diet. Although mice lacking both Sestrin1 and Sestrin2 only in adipose tissue (AdiDKO mice) lack mTORC1 regulation in the WAT in response to leucine (fig. S12, A and B), they did not phenocopy DKO mice when fed the leucine-free diet (fig. S12, C to E), indicating that Sestrin loss affects WAT mass through a tissue-nonautonomous mechanism.

Given that hepatic mTORC1 can regulate organismal physiology (*44*–*46*) and is sensitive to dietary leucine in a Sestrin-dependent manner (Fig. 1E and fig. S1A), we hypothesized a central role for the Sestrins in the liver in the response to leucine limitation. Indeed, in mice lacking both Sestrins in the liver (LiDKO mice), leucine deprivation caused a greater loss of body weight than in wild-type and *Sesn2*$^{-/-}$ mice, albeit not as pronounced as in DKO mice (Fig. 3A). As DKO mice lose both WAT and muscle mass upon leucine deprivation (Fig. 2, C to G), we sought to determine what accounted for the reduction in body weight of LiDKO mice. On the leucine-free diet, LiDKO mice phenocopied the severe WAT loss of DKO mice (Fig. 3, B to D) but had muscle mass and plasma amino acid concentrations similar to those in wild-type animals (Fig. 3E and fig. S13). The WAT loss of DKO and LiDKO mice is unlikely to be the consequence of an unrecognized developmental defect, as the AAV-Cre–mediated acute deletion of Sestrin1 and Sestrin2 in the livers of adult mice conferred the same phenotype (fig. S14).

Together, our results indicate that hepatic Sestrin-mTORC1 plays a key role in the organismal response to leucine deprivation and suggest that liver-to-WAT communication controls the WAT loss observed in DKO mice. The two other phenotypes we documented in DKO mice—loss of muscle mass and deregulation of plasma amino acid concentrations upon leucine starvation—are independent of Sestrin function in the liver and are perhaps mediated by the Sestrins in muscle, given the importance of muscle for maintaining circulating amino acid concentrations (*47*–*50*). Thus, mice require leucine-sensitive Sestrin function in several tissues to maintain homeostasis upon removal of dietary leucine.

Among its many functions, the liver is a source of circulating factors, or hepatokines, that promote metabolic homeostasis. Among these, we focused on fibroblast growth factor 21 (FGF21), as it is implicated in the response to amino acid starvation and WAT remodeling (*51*–*53*). Upon leucine deprivation, plasma FGF21 concentrations were higher in DKO than wild-type mice (Fig. 4A) but similar when mice were fed an amino acid–replete diet (Fig. 4A) or fasted (fig. S15A). In DKO mice, FGF21 mediates the exacerbated loss of body weight and WAT during leucine deprivation, as its

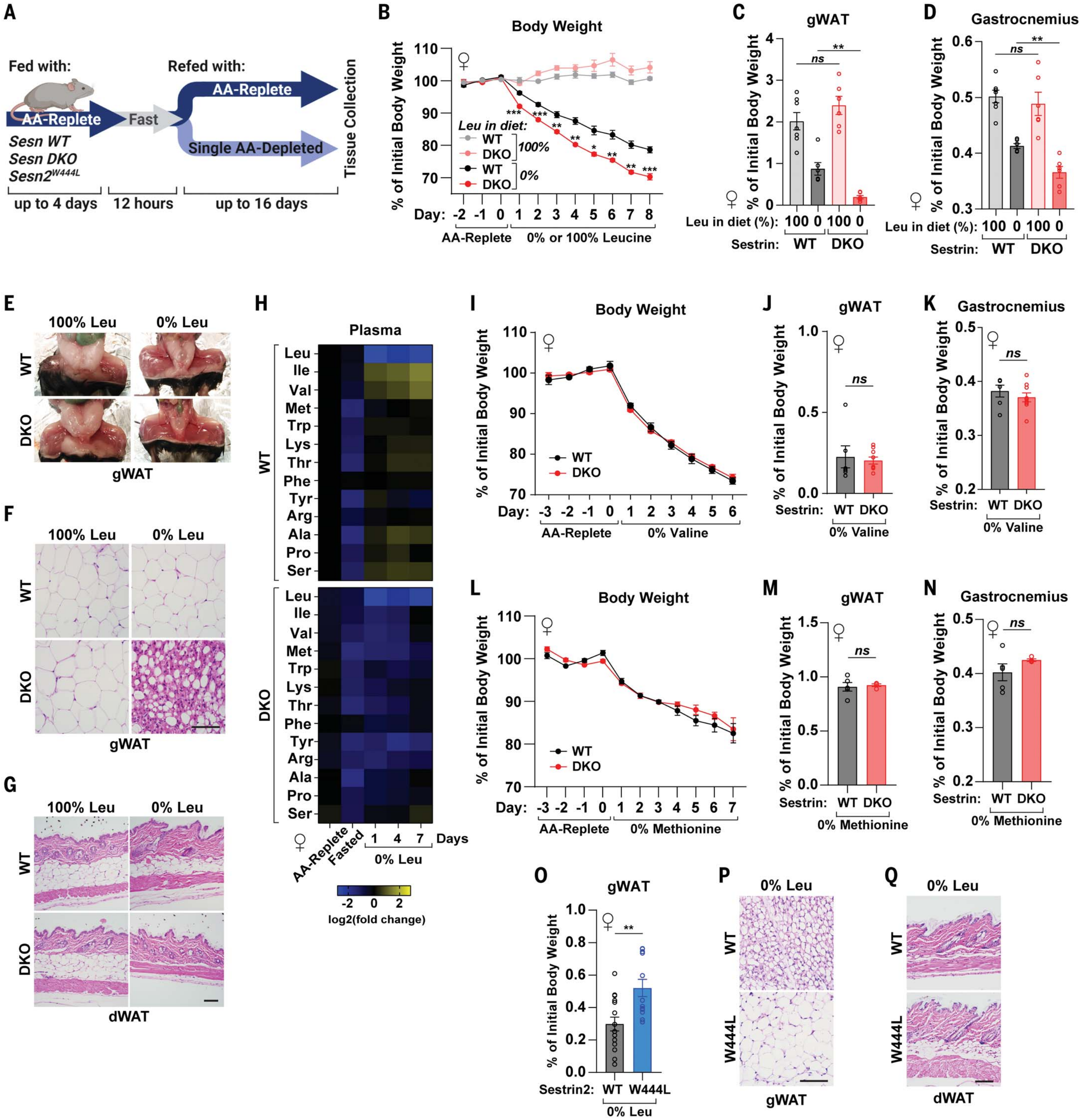


Fig. 2. Mice require Sestrin1 and Sestrin2 to adapt to limitations in dietary leucine. (**A**) Experimental setup for studying the long-term impacts of depriving mice of individual amino acids. Mice of the indicated genotypes were maintained on an amino acid–replete control diet for up to 4 days, fasted overnight for 12 hours, and then refed with the control diet or food lacking an essential amino acid for up to 16 days. (**B**) Body weights of female mice of the indicated genotypes during feeding with the indicated diets (n = 6 to 7 mice). The daily body weight measurements of each mouse during initial maintenance on an amino acid–replete control diet were averaged; the percent change from this average is depicted. (**C** and **D**) Gonadal WAT (C) and gastrocnemius muscle weights (D) in female mice of the indicated genotypes after 8 days on the indicated diets (n = 6 to 7 mice). Tissue weights for each mouse are presented as the percent of the average body weight while on an amino acid–replete control diet. (**E**) Representative images of gonadal WAT from female mice of the indicated genotypes after 8 days on the indicated diets (n = 6 to 7 mice). (**F** and **G**) Hematoxylin and eosin (H&E) stain of gonadal WAT (F) and dermal WAT (dWAT) (G) pad sections from female mice of the indicated genotypes after 8 days on the indicated diets. Images are representative of 6 to 7 mice. Scale bars, 50 μm. (**H**) Relative plasma abundances of amino acids from serial blood sampling of female mice of the indicated genotypes, which were kept on an amino acid–replete diet, fasted for 12 hours overnight, and then refed with a leucine-free diet for up to 7 days. Data are presented as $\log_2$ fold change of mean values relative to those in wild-type mice on the control diet (n = 3 to 5 mice). Amino acids with significant changes

($P < 0.05$) during leucine-free feeding as compared with the amino acid–replete condition are shown. See fig. S9 for all amino acids and statistical analyses. (**I**) Body weights of female mice of the indicated genotypes on a valine-free diet (n = 6 to 9 mice). The daily body weight measurements of each mouse during initial maintenance on an amino acid–replete control diet were averaged; the percent change from this average is depicted. (**J** and **K**) Gonadal WAT (J) and gastrocnemius muscle weight (K) of female mice of the indicated genotypes after 8 days of feeding on a valine-free diet (n = 6 to 9 mice). Tissue weight for each mouse is presented as percent of the average body weight while on the amino acid–replete control diet. (**L** to **N**) Same analyses as in (I) to (K), respectively, except mice were fed a methionine-free diet (n = 4 to 5 mice). (**O**) Gonadal WAT weight of female mice of the indicated genotypes after 16 days of feeding with a leucine-free diet (n = 12 to 15 mice). Tissue weight for each mouse is presented as percent of the average body weight while initially kept on the amino acid–replete control diet for 4 days. (**P** and **Q**) H&E stain of gonadal (P) and dermal (Q) WAT pad sections from female mice of the indicated genotypes after 16 days of feeding with a leucine-free diet. Images are representative of 6 to 8 mice. Scale bars, 50 μm. Data are the mean ± SEM. P values were determined using repeated measures two-way ANOVA with Sidak test [(B), (I), and (L)] or two-tailed t tests [(C), (D), (J), (K), and (M) to (O)]. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

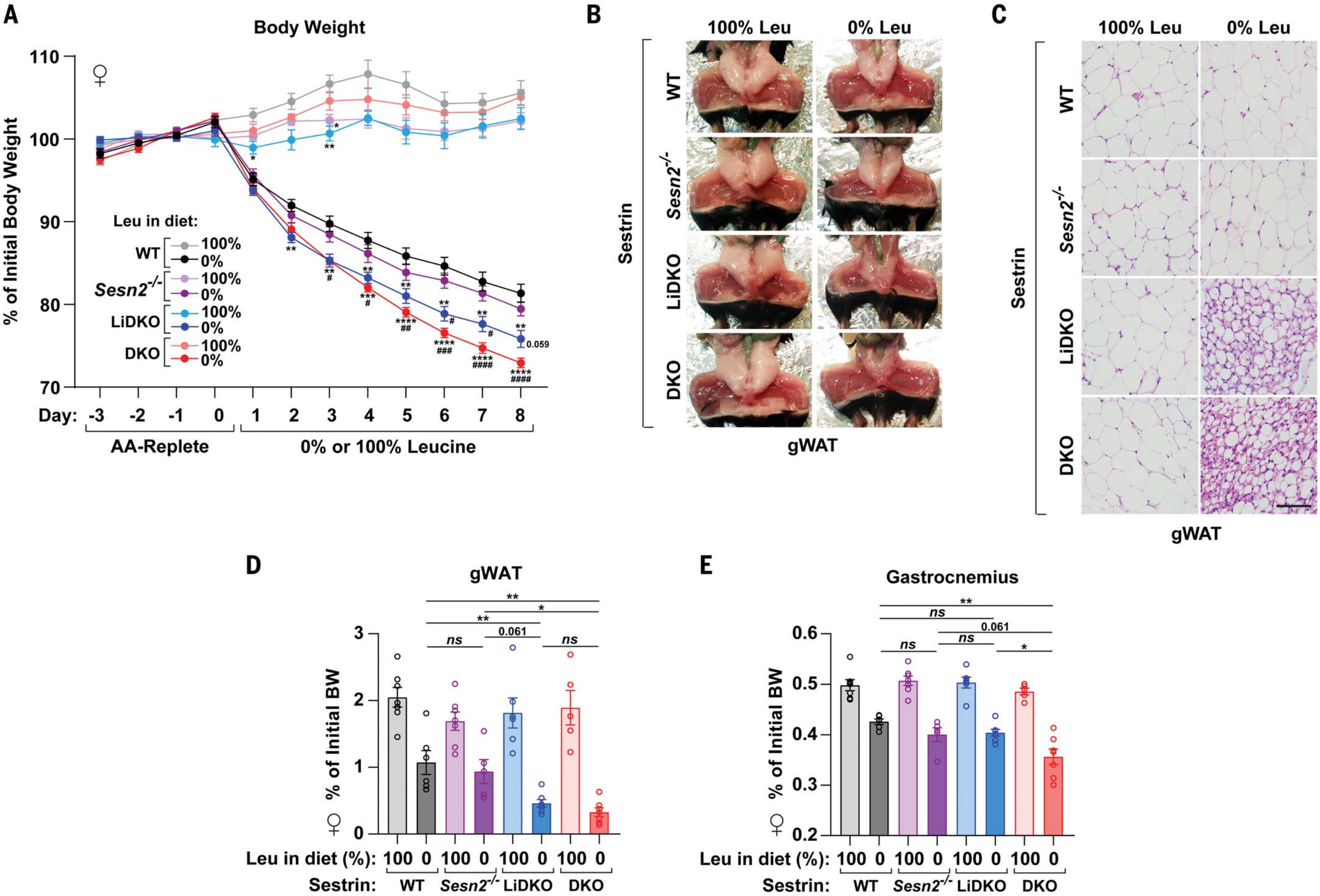


Fig. 3. Liver Sestrins control WAT remodeling upon deprivation of dietary leucine. (**A**) Body weights of female mice of the indicated genotypes fed the indicated diets (n = 5 to 12 mice). The daily body weight measurements of each mouse during initial maintenance on an amino acid–replete control diet were averaged; the percent change from this average is depicted. Statistical comparisons to the wild-type group (*) and the $Sesn2^{-/-}$ group (#) are shown. (**B**) Images of gonadal WAT in female mice of the indicated genotypes after 8 days on the indicated diets. Images are representative of 3 to 4 mice. (**C**) H&E analyses of gonadal WAT sections from female mice of the indicated genotypes after 8 days on the indicated diets. Images are representative of 3 to 4 mice. Scale bar, 50 μm. (**D** and **E**) Gonadal WAT (D) and gastrocnemius muscle (E) weights of female mice with the indicated genotypes after 8 days on a leucine-free diet (n = 5 to 7 mice). Tissue weights of each mouse are presented as percent of the average body weight (BW) on the amino acid–replete control diet. Data are the mean ± SEM. P values were determined using repeated measures two-way ANOVA with Tukey test (A) or one-way ANOVA with Tukey test [(D) and (E)]. *$P < 0.05$, **$P < 0.01$.

deletion prevented these effects (Fig. 4, B to E, and fig. S15, B to D). When deprived of leucine, DKO and LiDKO mice had similar increases in plasma FGF21 (Fig. 4A), pinpointing the liver as the likely source of the FGF21. Correspondingly, leucine deprivation increased the amount of *Fgf21* mRNA in the liver of DKO mice (fig. S15E), resulting in a boost in hepatic FGF21 protein amounts that strongly correlated with those in the plasma (fig. S15F). Consistent with estrogen signaling potentiating hepatic FGF21 production (*54*), FGF21

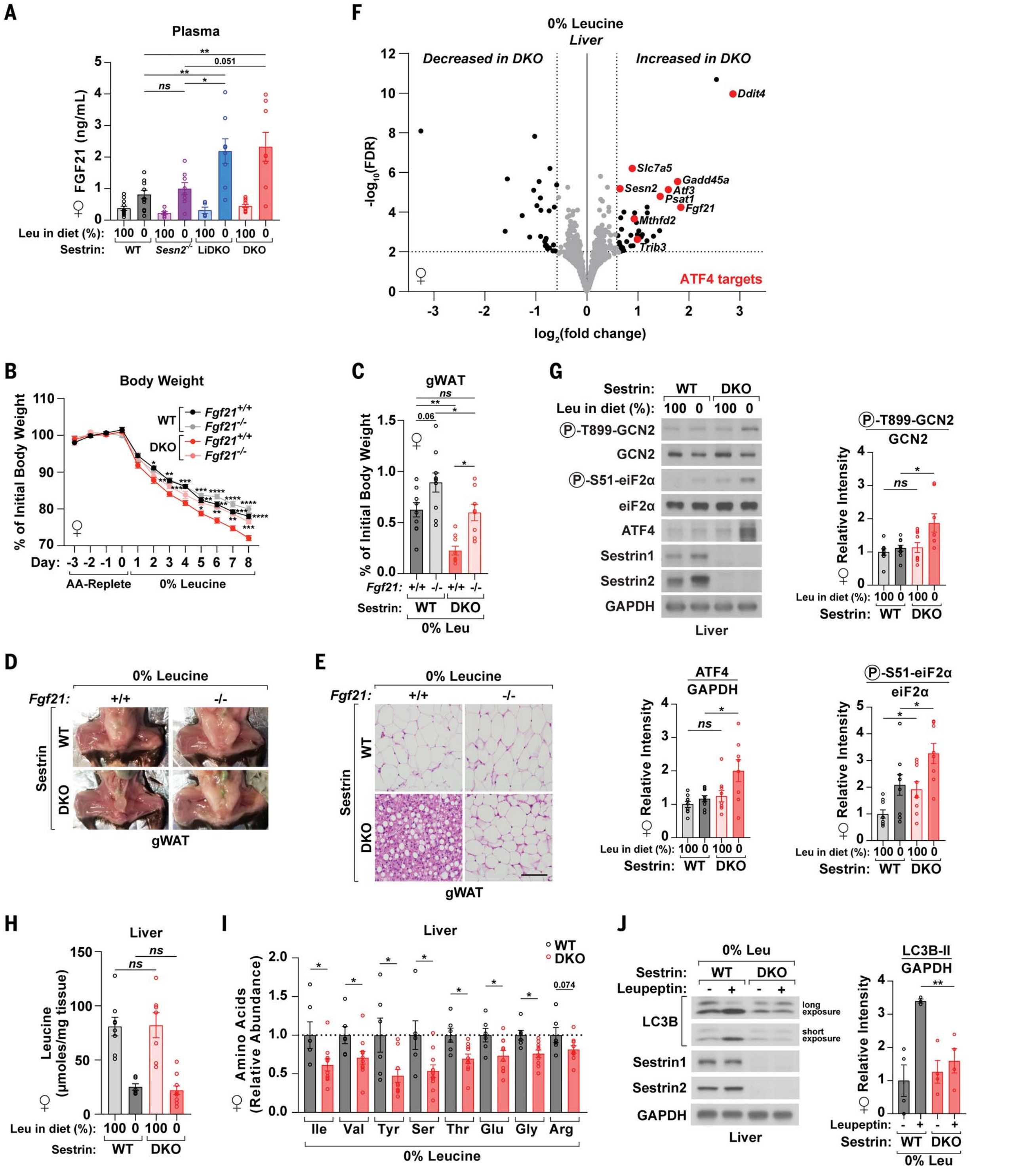


Fig. 4. During dietary leucine deprivation, Sestrins in the liver control WAT maintenance through FGF21 production. (**A**) Plasma FGF21 concentrations in female mice of the indicated genotypes 24 hours after feeding with the indicated diets (n = 5 to 12 mice). (**B**) Body weights of female mice of the indicated genotypes during feeding with a leucine-free diet (n = 5 to 10 mice). The daily body weight measurements of each mouse during initial maintenance on an amino acid–replete control diet were averaged; the percent change from this average is depicted. Statistical comparisons to the DKO group are shown. (**C**) Gonadal WAT weight of female mice with the indicated genotypes after 8 days on a leucine-free diet (n = 5 to 10 mice). Tissue weights of each mouse are presented as percent of the average body weight on the amino acid–replete control diet. (**D**) Images of gonadal WAT in female mice of the indicated genotypes after 8 days on a leucine-free diet. Images are representative of 5 to 10 mice. (**E**) H&E analyses of gonadal WAT sections from female mice of the indicated genotypes after 8 days on a leucine-free diet. Images are representative of 5 to 10 mice. Scale bar, 50 μm. (**F**) Volcano plot of genes differentially expressed in WT and DKO livers after 24 hours of leucine-free feeding (n = 7 to 18 mice). Transcripts that are differentially expressed ≥1.5-fold with a false discovery rate (FDR) of <0.01 are depicted in black. Among these, ATF4 target genes are depicted in red. For better visualization, *Sesn1* [$\log_2$(fold change) = −1.60;

$-\log_{10}$(FDR) = 79.87] was excluded from the plot. *Sesn2* reads in DKO mice are derived from nonfunctional transcripts generated by the *Sesn2* null allele. See fig. S16 for additional analysis. (**G**) Female mice of the indicated genotypes were fed with the indicated diets for 24 hours, and liver lysates were analyzed by immunoblotting for the phosphorylation state and amounts of the indicated proteins (n = 8 to 9 mice). (**H**) Quantification of leucine in the livers of female mice of the indicated genotypes after 24 hours of feeding with the indicated diets (n = 6 to 10 mice). Molar quantities are normalized to tissue weights. (**I**) Relative abundances of amino acids in the livers of female mice of the indicated genotypes after 24 hours of feeding with a leucine-free diet. Abundances are normalized to tissue weights and shown relative to the average abundances in wild-type livers (n = 6 to 10 mice). Data were acquired from the same samples as in (H). Amino acids with significant changes ($P < 0.05$) are shown. See fig. S20A for data for all amino acids and experimental groups. (**J**) Female mice of the indicated genotypes were treated with leupeptin or vehicle for 4 hours after 24 hours of feeding with a leucine-free diet. Liver lysates were analyzed by immunoblotting for amounts of the indicated proteins (n = 3 to 4 mice). Data are the mean ± SEM. P values were determined using one-way ANOVA with Tukey test [(A) and (C)], repeated measures two-way ANOVA with Tukey test (B), or two-tailed t test [(G) to (J)]. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

concentrations increased less in male than female DKO mice deprived of leucine (fig. S15G), perhaps accounting for the more modest loss of WAT observed in males (fig. S5C).

Fgf21 expression in the liver is transcriptionally regulated by several mechanisms, including the transcription factor ATF4 (*55–57*). Transcriptome analysis revealed that during leucine deprivation, DKO mice strongly induced ATF4 activity in the liver, as they had increased expression of many ATF4 target genes compared with leucine-deprived wild-type controls at short (24 hours) and long (8 days) timepoints after the start of the leucine-free diet (Fig. 4F and fig. S16). ATF4 is induced by eukaryotic translation initiation factor 2 subunit alpha (eIF2α) phosphorylation during the integrated stress response (ISR) (*58*). Consistent with activation of the ISR in the livers of DKO mice, leucine starvation led to an increase in eIF2α phosphorylation and ATF4 protein (Fig. 4G). Independently of eIF2α, mTORC1 itself can also promote *Atf4* mRNA translation (*59*, *60*) and so may contribute to the observed increase in ATF4 protein. Oxidative stress can also induce ATF4 (*61*, *62*), but we observed no differences in redox state between wild-type and DKO livers during leucine deprivation, as indicated by the ratios of GSH:GSSG, NADH:NAD+, and AMP:ATP (fig. S17, A to C).

Several stress-activated kinases can phosphorylate eIF2α, including PERK, which responds to endoplasmic reticulum (ER) stress, and GCN2, which is activated by the uncharged tRNAs that accumulate when amino acids are limiting for tRNA aminoacylation (*63*). Leucine starvation did not increase PERK autophosphorylation, a marker of its activity, as compared with treatment with tunicamycin, a canonical activator of ER stress (fig. S17D). In contrast, leucine deprivation strongly and rapidly boosted GCN2 activity (indicated by its autophosphorylation) and ATF4 target gene expression in the liver during the initial 3 to 6 hours of refeeding with the leucine-free diet (figs. S18 and S19). Notably, while wild-type mice largely showed suppression of this initial ISR induction within 24 hours of leucine starvation, DKO mice retained elevated GCN2 activity (Fig. 4G), suggesting that the Sestrins are required for the liver to maintain amino acid homeostasis during leucine deprivation. Indeed, while leucine starvation reduced leucine to equally low levels in the livers of both wild-type and DKO mice (Fig. 4H and fig. S20A), wild-type mice largely maintained or even increased the levels of other proteogenic amino acids in the liver, but DKO mice did not and had lower hepatic levels of several amino acids (Fig. 4I and fig. S20A).

A major source of amino acids in the liver is the degradation of proteins by autophagy (*64*, *65*), a process suppressed by mTORC1 through inhibitory phosphorylation of the kinase ULK1 (*4*, *21–23*). Notably, many of the amino acids reduced in the livers of DKO mice, particularly isoleucine, valine, threonine, tyrosine, and serine, are also affected in the livers of mice with hepatic autophagy disruption (*65*). We therefore examined autophagic flux by measuring LC3B lipidation after treating mice with the lysosomal protease inhibitor leupeptin (*66*, *67*). Compared with a control diet, leucine deprivation increased autophagic flux in wild-type livers (fig. S20B). In contrast, in DKO mice, leucine deprivation did not relieve the inhibitory phosphorylation of ULK1 by mTORC1 (fig. S20C) or induce autophagic flux (Fig. 4J), providing a likely explanation for their activation of GCN2 and expression of ATF4 and FGF21.

Protein synthesis can also affect levels of amino acids through their consumption and is regulated by both mTORC1 and GCN2 (*68*). However, we found no significant impact of Sestrin expression on protein synthesis in the liver, as assessed by polysome profiling and puromycin incorporation into protein (fig. S21).

In DKO mice on the leucine-free diet, inhibition of mTORC1 with rapamycin attenuated the aberrant phosphorylation of mTORC1 substrates, drop in hepatic amino acids, ATF4 target gene expression, and plasma FGF21 concentrations (fig. S22). Together, these results confirm that, as with WAT mass (fig. S10, B to D), the Sestrins affect leucine-sensitive physiology through mTORC1.

Leucine sensing is spatially compartmentalized in the liver and drives a zonated response to dietary leucine deprivation

Within the liver, hepatocytes are organized into a large number of hexagonally shaped lobules. Nutrient-rich blood coming from the gastrointestinal tract enters the periphery of each lobule through branches of the portal vein and percolates through sinusoids between the hepatocytes before exiting at the central vein. Hepatocytes in different zones of the lobule can have distinct metabolic functions and transcriptional programs (*69*, *70*), so we reasoned that there might be a spatial organization to Sestrin1 and Sestrin2 expression within the liver lobule. Indeed, using single-molecule fluorescence in situ hybridization, we observed zonated expression of the *Sesn1* and *Sesn2* mRNAs, with many transcripts in periportal and midlobular hepatocytes and fewer in the pericentral hepatocytes marked by *Glul* expression (Fig. 5A and fig. S23, A to C). This finding suggested that there also might be zonal differences in the leucine sensitivity of mTORC1 signaling, which we monitored using an immunofluorescence assay for phosphorylated S6, a marker of mTORC1 activity. Notably, mTORC1 activity was not uniform across the liver lobule in wild-type mice fed a leucine-free diet. It was inhibited in the periportal and midlobular hepatocytes that express Sestrin but active in the pericentral ones that do not. In contrast, in DKO mice starved of leucine, mTORC1 activity was uniform across the lobule and indistinguishable from that in mice fed amino acid–replete food (Fig. 5B and fig. S23D).

To understand the impact of zonated mTORC1 activity, we determined the spatial pattern of *Fgf21* expression in wild-type and DKO mice. *Fgf21* was minimally expressed in the livers of mice fed a control diet but was induced upon leucine deprivation in the periportal and midlobular hepatocytes that express Sestrin. This induction was amplified by Sestrin loss (Fig. 5C and fig. S23E), consistent

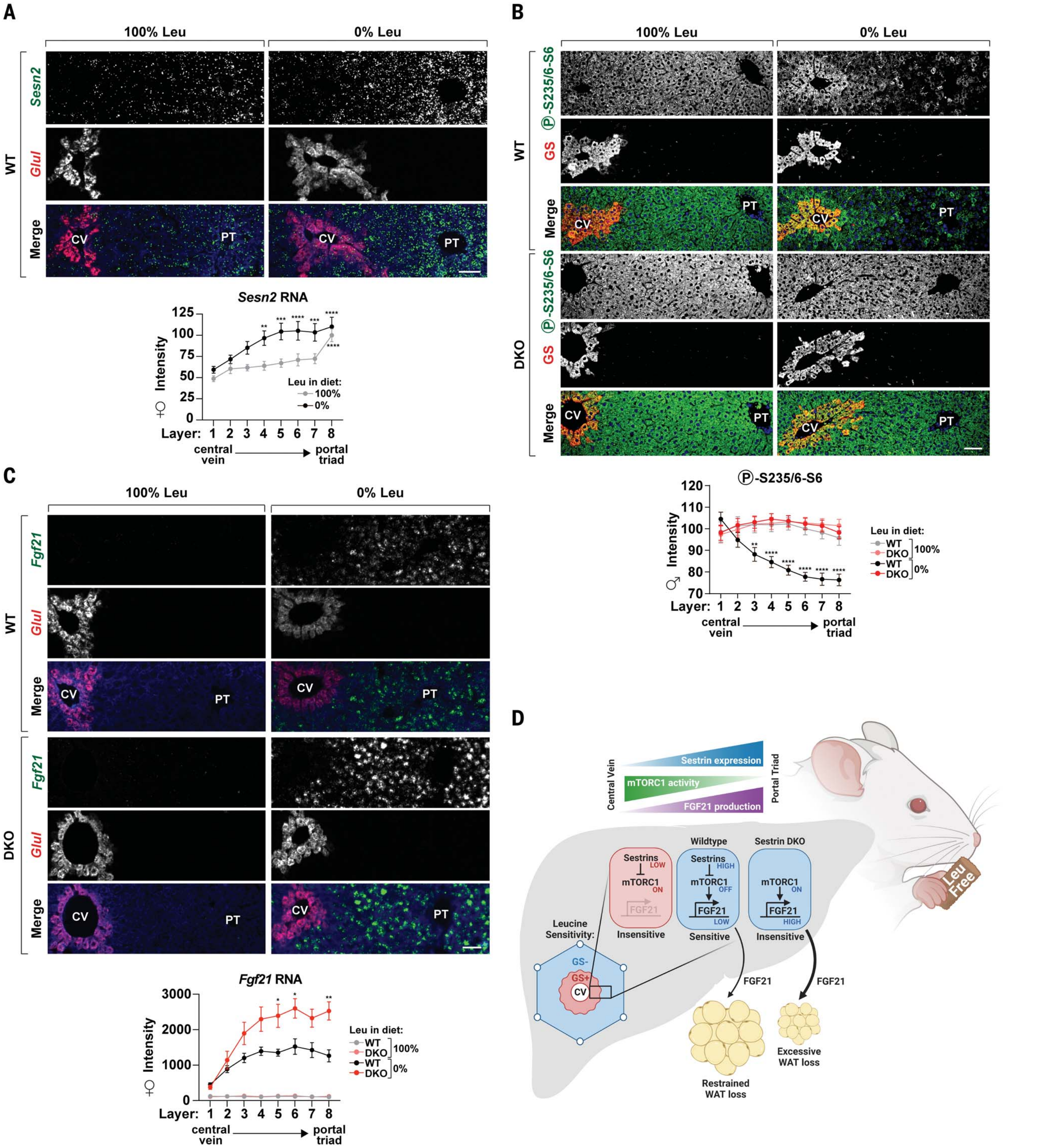


Fig. 5. Zonated Sestrin expression establishes leucine-sensitive and leucine-insensitive compartments in the liver. (**A**) Representative images and quantification of *Sesn2* mRNA in the livers of wild-type female mice 24 hours after feeding with the indicated diets (16 to 18 lobules from three mice per diet). *Glul* encoding glutamine synthetase marks pericentral hepatocytes. Shown are statistical comparisons to layer 1 for each diet. (**B**) Representative images and quantification of S6 phosphorylation as detected in immunofluorescence assays in liver sections from female mice of the indicated liver genotypes 3 hours after refeeding with the indicated diets (12 to 24 lobules from four to seven mice per genotype per diet). GS indicates glutamine synthetase and marks pericentral hepatocytes. Statistical comparisons between genotypes on the leucine-free diet are shown. (**C**) Representative images and quantification of *Fgf21* mRNA in liver sections of female mice of the indicated genotypes after 24 hours of feeding with the indicated diets (7 to 16 lobules from three mice per genotype per diet). *Glul* encoding glutamine synthetase marks pericentral hepatocytes. Statistical comparisons between genotypes on the leucine-free diet are shown. (**D**) Model of zonated leucine sensing by Sestrin-mTORC1 in the liver and its role in the physiological response to dietary leucine deprivation. CV, central vein; PT, portal triad. Scale bars, 50 μm. Data are the mean ± SEM. *P* values were determined using two-way ANOVA with Dunnett's test (A) or two-way ANOVA with Sidak test [(B) and (C)]. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

with the increase in ATF4 detected by immunoblotting (Fig. 4G). Furthermore, *Sesn2*, which is itself an ATF4 target (*71*), was also induced by leucine deprivation in wild-type mice (Fig. 4G) in the same zones as *Fgf21* (Fig. 5, A and C). The pericentral hepatocytes did not activate the ISR during leucine deprivation, as indicated by their lack of *Fgf21* induction, despite maintaining high mTORC1 activity. This suggests that they are intrinsically resistant to the negative effects of leucine deprivation, perhaps accounting for why they are wired to have low Sestrin expression. The lack of ISR activation in these cells may itself contribute to their low expression of *Sesn2*, as it is an ATF4 target gene. The resistance of the pericentral hepatocytes may stem from known features of these cells, including their constitutively high levels of autophagy (*72*, *73*), expression of ER chaperones (*74*), and synthesis of intracellular glutamine whose efflux can drive leucine uptake (*75*). We conclude that in the liver, the Sestrin-imposed and zone-specific regulation by leucine of mTORC1 signaling is necessary to attenuate the stress of dietary leucine deprivation (Fig. 5D).

Discussion

The nature of the leucine sensing pathway upstream of mTORC1 has been controversial, and many diverse sensors and mechanisms have been proposed to play a role (*24*–*31*), largely on the basis of work in cultured cells. We find that Sestrin1 and Sestrin2 control mTORC1 activity in response to leucine in vivo and that this modulation is necessary for mice to adapt to limitations in dietary leucine. These data, along with previous biochemical and structural work (*32*, *33*), are consistent with Sestrin1 and Sestrin2 being leucine sensors for the mTORC1 pathway. It has been proposed that the Sestrins can function through AMPK (*31*), but we found that AMPK is not required for dietary leucine to regulate mTORC1 in our model system. This does not preclude the possibility, however, that Sestrin loss may secondarily affect AMPK, because mTORC1 activity is known to drive energetic stress (*76*).

Further, our results suggest a temporal relationship between the activities of mTORC1 and GCN2 in the response to dietary leucine deficiency, in which Sestrin-mediated mTORC1 inhibition is necessary to maintain amino acid homeostasis and attenuate GCN2 activity during prolonged, but not acute, leucine deprivation. As GCN2 also modulates mTORC1 activity through ATF4-mediated *Sesn2* expression (*71*), our data reflect a homeostatic mechanism of reciprocal cross-talk between these amino acid–sensitive kinases.

We also reveal a previously unappreciated complexity to nutrient sensing by mTORC1 in vivo. In the liver lobule, only hepatocytes in certain zones appreciably express the Sestrins, and only in these zones is mTORC1 signaling sensitive to dietary leucine. In considering why this might be the case, it is worth recalling that mTORC1 controls a large number of metabolic pathways and is also regulated by a diverse set of nutrients (*4*, *21*–*23*). Such an arrangement sets up a conundrum: How can mTORC1 regulate a metabolic pathway in response to the concentrations of a specific nutrient—such as one consumed by the pathway—but not to others to which sensitivity would be unfavorable?

We hypothesize that, at least in the liver, the zonated expression of nutrient sensors is part of the answer, along with the well-appreciated zonation of many metabolic processes. Although at the whole-liver level we were unable to detect changes in protein synthesis, it is noteworthy that the periportal hepatocytes that express the Sestrins also have been reported to play a predominant role in liver protein synthesis (*77*, *78*), a process that can be regulated by mTORC1 and consumes leucine. Conversely, the pericentral hepatocytes that have low Sestrin expression are the main site of ketogenesis, a process that mTORC1 can inhibit (*47*) but for which leucine sensitivity would be inappropriate given that an abundant supply of dietary glucose, regardless of leucine availability, renders ketones redundant as a fuel source. The combinatorial impact of the zonated expression of the nutrient sensors that signal to mTORC1 and of the metabolic processes controlled by it may underlie how one pathway can appropriately regulate such a variety of metabolic processes in response to diverse nutritional states.

REFERENCES AND NOTES

1. M. Wallace *et al.*, *Nat. Chem. Biol.* **14**, 1021–1031 (2018).
2. S. B. Crown, N. Marze, M. R. Antoniewicz, *PLOS ONE* **10**, e0145850 (2015).
3. J. Rosenthal, A. Angel, J. Farkas, *Am. J. Physiol.* **226**, 411–418 (1974).
4. R. A. Saxton, D. M. Sabatini, *Cell* **168**, 960–976 (2017).
5. J. C. Anthony *et al.*, *J. Nutr.* **130**, 2413–2419 (2000).
6. Y. Duan *et al.*, *Front. Biosci. (Landmark Ed.)* **20**, 796–813 (2015).
7. Y. Yin *et al.*, *Amino Acids* **39**, 1477–1486 (2010).
8. F. Li, Y. Yin, B. Tan, X. Kong, G. Wu, *Amino Acids* **41**, 1185–1193 (2011).
9. W. T. Moore *et al.*, *Curr. Diab. Rep.* **15**, 76 (2015).
10. C. A. de Oliveira, M. Q. Latorraca, M. A. de Mello, E. M. Carneiro, *Amino Acids* **40**, 1027–1034 (2011).
11. J. Yang, Y. Chi, B. R. Burkhardt, Y. Guan, B. A. Wolf, *Nutr. Rev.* **68**, 270–279 (2010).
12. E. A. Ananieva, J. D. Powell, S. M. Hutson, *Adv. Nutr.* **7**, 798S–805S (2016).
13. W. Ren *et al.*, *Cell Death Dis.* **8**, e2655 (2017).
14. M. Torigoe *et al.*, *Mod. Rheumatol.* **29**, 885–891 (2019).
15. N. E. Richardson *et al.*, *Nat. Aging* **1**, 73–86 (2021).
16. G. D'Antona *et al.*, *Cell Metab.* **12**, 362–372 (2010).
17. S. M. Solon-Biet *et al.*, *Nat. Metab.* **1**, 532–545 (2019).
18. C. B. Newgard *et al.*, *Cell Metab.* **9**, 311–326 (2009).
19. C. J. Lynch, S. H. Adams, *Nat. Rev. Endocrinol.* **10**, 723–736 (2014).
20. L. Fontana *et al.*, *Cell Rep.* **16**, 520–530 (2016).
21. G. Y. Liu, D. M. Sabatini, *Nat. Rev. Mol. Cell Biol.* **21**, 183–203 (2020).
22. K. J. Condon, D. M. Sabatini, *J. Cell Sci.* **132**, jcs222570 (2019).
23. A. J. Valvezan, B. D. Manning, *Nat. Metab.* **1**, 321–333 (2019).
24. J. M. Han *et al.*, *Cell* **149**, 410–424 (2012).
25. J. H. Kim *et al.*, *Nat. Commun.* **8**, 732 (2017).
26. X. D. He *et al.*, *Cell Metab.* **27**, 151–166.e6 (2018).
27. R. V. Durán *et al.*, *Mol. Cell* **47**, 349–358 (2012).
28. J. F. Linares *et al.*, *Mol. Cell* **51**, 283–296 (2013).
29. R. E. Lawrence *et al.*, *Nat. Cell Biol.* **20**, 1052–1063 (2018).
30. S. M. Son *et al.*, *Cell Metab.* **29**, 192–201.e7 (2019).
31. A. V. Budanov, M. Karin, *Cell* **134**, 451–460 (2008).
32. R. L. Wolfson *et al.*, *Science* **351**, 43–48 (2016).
33. R. A. Saxton *et al.*, *Science* **351**, 53–58 (2016).
34. Z. Fang *et al.*, *Cell. Mol. Gastroenterol. Hepatol.* **12**, 921–942 (2021).
35. B. A. Yang *et al.*, *Stem Cell Reports* **16**, 2078–2088 (2021).
36. J. H. Lee *et al.*, *Cell Metab.* **16**, 311–321 (2012).
37. J. Segalés *et al.*, *Nat. Commun.* **11**, 189 (2020).
38. M. Kim *et al.*, *Nat. Commun.* **11**, 190 (2020).
39. J. Lu *et al.*, *Nat. Aging* **1**, 60–72 (2021).
40. F. Xiao *et al.*, *Diabetes* **60**, 746–756 (2011).
41. S. Wei *et al.*, *Heliyon* **4**, e00830 (2018).
42. D. Xu *et al.*, *Am. J. Physiol. Endocrinol. Metab.* **316**, E817–E828 (2019).
43. J. Magdalon *et al.*, *Biochim. Biophys. Acta* **1861**, 430–438 (2016).
44. S. Sengupta, T. R. Peterson, M. Laplante, S. Oh, D. M. Sabatini, *Nature* **468**, 1100–1104 (2010).
45. Y. Koketsu *et al.*, *Am. J. Physiol. Endocrinol. Metab.* **294**, E719–E725 (2008).
46. M. Cornu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11592–11599 (2014).
47. T. Pozefsky, R. G. Tancredi, R. T. Moxley, J. Dupre, J. D. Tobin, *J. Clin. Invest.* **57**, 444–449 (1976).
48. M. H. Vendelbo *et al.*, *PLOS ONE* **9**, e102031 (2014).
49. G. F. Cahill Jr., *N. Engl. J. Med.* **282**, 668–675 (1970).
50. R. R. Wolfe, *Am. J. Clin. Nutr.* **84**, 475–482 (2006).
51. A. L. De Sousa-Coelho, P. F. Marrero, D. Haro, *Biochem. J.* **443**, 165–171 (2012).
52. A. L. De Sousa-Coelho *et al.*, *J. Lipid Res.* **54**, 1786–1797 (2013).
53. T. Laeger *et al.*, *J. Clin. Invest.* **124**, 3913–3922 (2014).
54. C. Allard *et al.*, *Mol. Metab.* **22**, 62–70 (2019).
55. T. Lundåsen *et al.*, *Biochem. Biophys. Res. Commun.* **360**, 437–440 (2007).
56. K. H. Kim *et al.*, *Nat. Med.* **19**, 83–92 (2013).
57. R. Maruyama, M. Shimizu, J. Li, J. Inoue, R. Sato, *Biosci. Biotechnol. Biochem.* **80**, 929–934 (2016).
58. H. P. Harding *et al.*, *Mol. Cell* **6**, 1099–1108 (2000).
59. M. E. Torrence *et al.*, *eLife* **10**, e63326 (2021).
60. I. Ben-Sahra, G. Hoxhaj, S. J. H. Ricoult, J. M. Asara, B. D. Manning, *Science* **351**, 728–733 (2016).
61. P. S. Lange *et al.*, *J. Exp. Med.* **205**, 1227–1242 (2008).
62. N. Miyamoto *et al.*, *Invest. Ophthalmol. Vis. Sci.* **52**, 1226–1234 (2011).
63. R. C. Wek, H. Y. Jiang, T. G. Anthony, *Biochem. Soc. Trans.* **34**, 7–11 (2006).
64. J. Madrigal-Matute, A. M. Cuervo, *Gastroenterology* **150**, 328–339 (2016).
65. J. Ezaki *et al.*, *Autophagy* **7**, 727–736 (2011).
66. M. Moulis, C. Vindis, *Cells* **6**, 14 (2017).
67. J. Haspel *et al.*, *Autophagy* **7**, 629–642 (2011).
68. T. G. Anthony *et al.*, *J. Biol. Chem.* **279**, 36553–36561 (2004).
69. K. B. Halpern *et al.*, *Nature* **542**, 352–356 (2017).
70. K. Jungermann, N. Katz, *Physiol. Rev.* **69**, 708–764 (1989).
71. J. Ye *et al.*, *Genes Dev.* **29**, 2331–2336 (2015).
72. R. Gebhardt, A. Hovhannisyan, *Dev. Dyn.* **239**, 45–55 (2010).
73. R. Gebhardt, P. J. Coffer, *Cell Commun. Signal.* **11**, 21 (2013).
74. C. Droin *et al.*, *Nat. Metab.* **3**, 43–58 (2021).
75. P. Nicklin *et al.*, *Cell* **136**, 521–534 (2009).
76. V. Aguilar *et al.*, *Cell Metab.* **5**, 476–487 (2007).
77. R. Gebhardt, M. Matz-Soja, *World J. Gastroenterol.* **20**, 8491–8504 (2014).
78. Y. Uchiyama, A. Asari, *Cell Tissue Res.* **236**, 305–315 (1984).

ACKNOWLEDGMENTS

We thank all members of the Sabatini lab, as well as H. Lodish and M. Vander Heiden, for suggestions and experimental help; B. Manning, T. Zhang, M. Wallace, C. Metallo, S. M. Jung, and D. Guertin for discussions; and M. Mihaylova, W. Festuccia, M. Abu-Remaileh, N. Laqtom, K. Lopez, and K. Knouse for technical advice and assistance. We thank M. Li for the *Sesn1*$^{flox/flox}$ and *Sesn2*$^{-/-}$ mice and the Gene Targeting and Transgenic Facility at Janelia Research Campus for generation of the *Sesn2*^{W444L} mice. We thank the MIT BioMicroCenter for RNA library prep and

sequencing. We also thank members of the Hope Babette Tang Histology Facility at the Koch Institute and S. Holder for histology support. Figures 1A, 2A, and 5D were created using Biorender.com. **Funding:** This work was funded by grants from the NIH (R01CA103866, R01CA129105, and R37AI047389) to D.M.S.; fellowship support from the NIH to A.L.C. (5F31DK113665), J.M.R. (F31CA232355), and J.B.S. (K00CA234839); a William N. and Bernice E. Bumpus Foundation Fellowship to A.M.P.; a Marie-Curie H2020 MSCA Global Fellowship (101033310) to A.A. D.M.S. is formerly an investigator of the Howard Hughes Medical Institute and an American Cancer Society research professor. **Author contributions:** A.L.C. and D.M.S. conceived of the project. A.L.C. designed and performed all experiments, with input from D.M.S. and assistance from A.M.P., T.P.N., J.M.R., A.A., and J.B.S., and A.M.P. helped with discussion and interpretation of results. T.K., B.W., S.H.C., and C.A.L. extracted metabolites, operated the liquid chromatography–mass spectrometry platform, and analyzed the metabolomics data. G.W.B. analyzed the RNA sequencing data. W.C.C. provided technical training, established the Sestrin knockout mouse lines, and designed the genetic strategy for generating the *Sesn2*^{W444L} mice. A.H. helped with discussion and execution of liver zonation experiments. A.L.C. and D.M.S. wrote the manuscript, and all authors edited it. **Competing interests:** D.M.S. is a shareholder of Navitor Pharmaceuticals, which is targeting the mTORC1 pathway for therapeutic benefit. **Data and materials availability:** All data are available in the main text or the supplementary materials. *Sesn2*^{W444L} mice will be deposited at a commercial animal vendor and made publicly available. The Gene Expression Omnibus accession number for the RNA sequencing data reported in this paper is GSE197806. To ensure sustainable access to data and materials associated with this study, the Whitehead Institute has committed to assuring long-term access and has designated the administrative email address sabadmin@wi.mit.edu as a contact point. Access to reagents will be facilitated by sabadmin@wi.mit.edu. Scientific inquiries can be sent to dmsabatini.lab@gmail.com. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9547
Materials and Methods
Figs. S1 to S23
Tables S1 and S2
References (*79*–*83*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 11 April 2021; resubmitted 2 March 2022
Accepted 24 May 2022
10.1126/science.abi9547

NEUROGENOMIC IMAGING

Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH

Rongxin Fang[1]†, Chenglong Xia[1]†‡, Jennie L. Close[2], Meng Zhang[1], Jiang He[1]§, Zhengkai Huang[1], Aaron R. Halpern[1], Brian Long[2], Jeremy A. Miller[2], Ed S. Lein[2], Xiaowei Zhuang[1]*

The human cerebral cortex has tremendous cellular diversity. How different cell types are organized in the human cortex and how cellular organization varies across species remain unclear. In this study, we performed spatially resolved single-cell profiling of 4000 genes using multiplexed error-robust fluorescence in situ hybridization (MERFISH), identified more than 100 transcriptionally distinct cell populations, and generated a molecularly defined and spatially resolved cell atlas of the human middle and superior temporal gyrus. We further explored cell-cell interactions arising from soma contact or proximity in a cell type–specific manner. Comparison of the human and mouse cortices showed conservation in the laminar organization of cells and differences in somatic interactions across species. Our data revealed human-specific cell-cell proximity patterns and a markedly increased enrichment for interactions between neurons and non-neuronal cells in the human cortex.

The human cerebral cortex comprises billions of cells of distinct types (*1*). The spatial organizations and interactions of these cells play a critical role in shaping and maintaining various brain functions (*2*). For instance, interactions between neuronal and non-neuronal cells are essential for axonal conduction, synaptic transmission, and tissue homeostasis and are required for normal functioning of the brain (*3*, *4*). Disruption of such cell-cell interactions contributes to various neurological disorders, such as autism (*5*), schizophrenia (*6*), and Alzheimer's disease (*7*). Yet we have only a limited understanding of the organizations and interactions of different cell types in the human cortex.

Recent single-cell RNA sequencing (scRNA-seq) analysis has revealed a diversity of transcriptionally distinct cell populations in the human middle temporal gyrus (MTG) (*8*). Efforts that combine scRNA-seq with microdissection (*8*), and more recently in situ sequencing to target 120 genes (*9*), have revealed the laminar organization of these transcriptionally defined neuronal cell types—in particular, the excitatory neurons—in the human MTG. These studies, however, did not elucidate the spatial relationship between cell types at high resolution, and a systematic characterization of cell-cell interactions among this high diversity of cell types is still lacking. Single-cell transcriptomics and epigenomics analyses have also provided rich insights into the evolution of cellular diversity and molecular signatures of cell types in the cortices of mice, marmosets, and humans (*8*, *10*, *11*), but how the spatial relationship and interactions between different cell types vary across species remains largely unclear.

Single-cell transcriptome imaging of the human cortex

Single-cell transcriptome imaging allows in situ gene expression profiling of individual cells and, hence, high-resolution spatial mapping of cell type organization in complex tissues. Here we describe single-cell transcriptome imaging of the human brain performed with multiplexed error-robust fluorescence in situ hybridization (MERFISH) (*12*). We carried out MERFISH measurements of the human MTG and superior temporal gyrus (STG) from fresh-frozen neurosurgical and postmortem brain samples, targeting 4000 genes (Fig. 1A). These genes included 764 differentially expressed marker genes in cell clusters derived from single-nucleus SMART-seq data of the MTG (*8*) and additional expressed genes that were largely randomly selected to increase the gene coverage. This allowed us to include potential marker genes not identified in the SMART-seq data, as well as functionally important genes such as ligands and receptors. To overcome the high autofluorescence background in human tissues due to lipofuscin, we photobleached the samples with light-emitting diode arrays (*13*) before MERFISH imaging. We then used expansion microscopy (*14*) to reduce the molecular crowding associated with imaging a large number of genes (*15*, *16*).

Individual RNA molecules were identified and assigned to segmented cells to determine the single-cell expression profiles (Fig. 1B and fig. S1). We imaged five tissue sections from neurosurgical MTG samples (from two male individuals, 36 and 32 years old) and five sections from postmortem STG samples (from two male individuals, 29 and 42 years old). MERFISH expression data showed excellent reproducibility between replicates (fig. S2, A and B); high correlation between neurosurgical MTG and postmortem STG samples, albeit with a lower total transcript count in the latter, likely due to RNA degradation (fig. S2C); and high correlation with bulk RNA sequencing data (fig. S3).

To test whether the molecular crowding associated with imaging 4000 genes caused substantial reduction in the detection efficiency, we performed MERFISH imaging on 250 of the 764 marker genes (in two expanded tissue sections for detection efficiency assessment and in three additional unexpanded sections to increase the number of cells imaged). The detection efficiency of the 4000-gene measurements was, on average, ~57% of that of the 250-gene measurements on expanded sections, with high correlation between the two measurements (fig. S4).

Cell type classification of the human cortex

We used single-cell expression profiles derived from the 4000-gene MERFISH data to identify transcriptionally distinct cell populations. First-level clustering detected excitatory and inhibitory neurons, as well as major subclasses of non-neuronal cells such as microglia, astrocytes, oligodendrocytes, oligodendrocyte progenitor cells (OPCs), endothelial cells, and mural cells, as characterized by the marker genes identified by SMART-seq (*8*) (fig. S5).

We then performed separate clustering analyses of inhibitory and excitatory neurons from the MERFISH data, which corresponded closely with those independently determined from the SMART-seq data (fig. S6). To combine information from both datasets, we performed integrated analysis of MERFISH and SMART-seq data (fig. S7, A and B). This analysis classified inhibitory neurons into four subclasses (denoted by marker genes *SST*, *VIP*, *PVALB*, and *LAMP5*, respectively) and excitatory neurons into nine subclasses (L2/3 IT, L4/5 IT, L5 IT, L6 IT, L6 IT *CAR3*, L5 ET, L5/6 NP, L6 CT, and L6b), with most subclasses further subdivided into multiple clusters (Fig. 1C). Because non-neuronal cells were depleted from the SMART-seq dataset (*8*), we identified clusters within individual subclasses of non-neuronal cells from the 4000-gene MERFISH data alone (Fig. 1C and fig. S7C). Altogether, we identified a total of 125 transcriptionally distinct cell populations in the human MTG and STG—29 excitatory, 39 inhibitory, and 57 non-neuronal clusters (Fig. 1C and fig. S7)—revealing not only a high diversity of neurons but also a high diversity of non-neuronal cells in the human cortex. To include the 250-gene data for downstream analysis, we performed supervised classification to predict their cell type labels (at the cluster level for neurons and the subclass level for non-neuronal cells) on the basis of annotations from the 4000-gene data.

[1]Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]Allen Institute for Brain Science, Seattle, WA 98109, USA.
*Corresponding author. Email: zhuang@chemistry.harvard.edu
†These authors contributed equally to this work.
‡Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.
§Present address: Vizgen, Inc., 61 Moulton Street, Cambridge, MA 02138, USA.

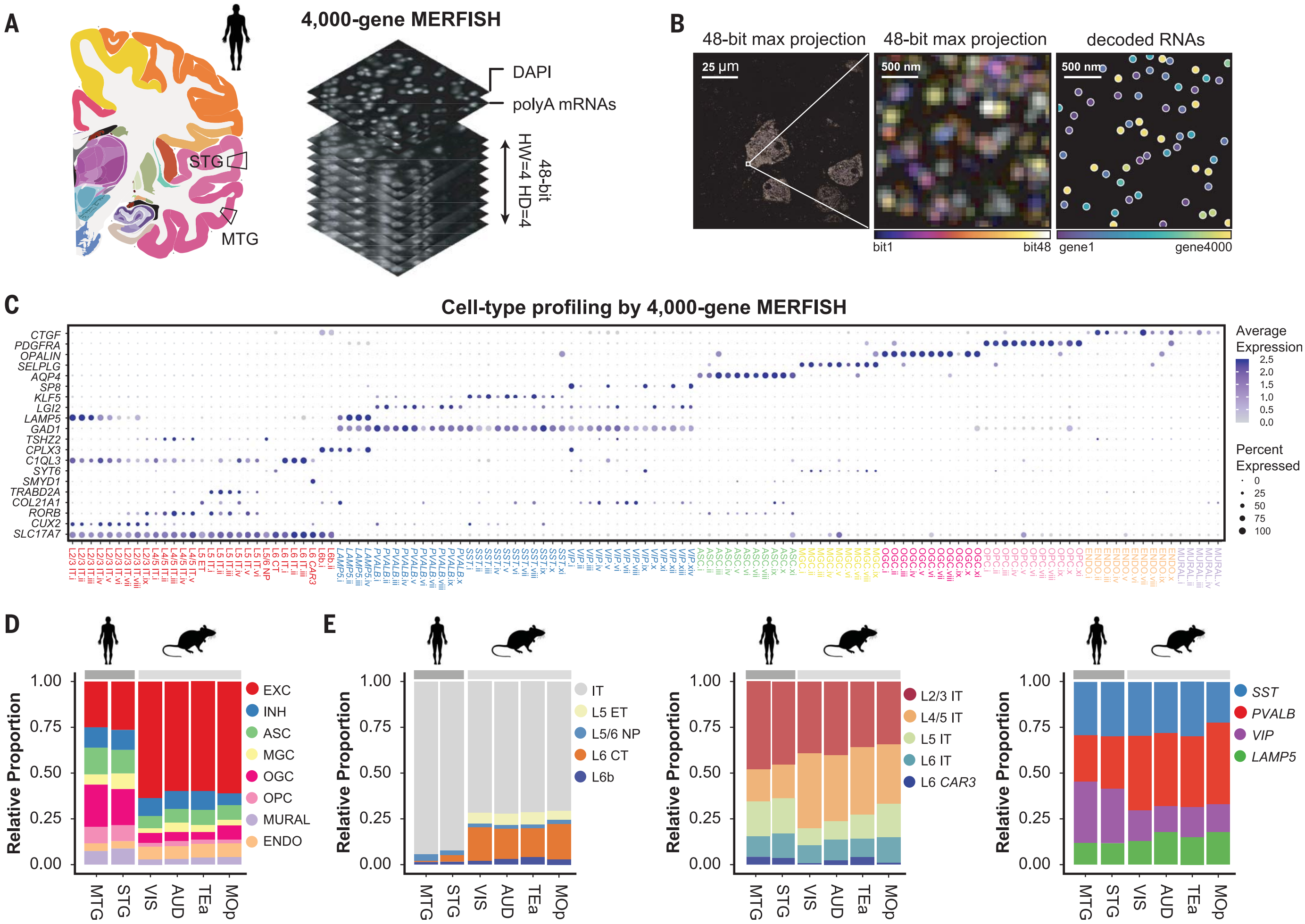


Fig. 1. Spatially resolved single-cell transcriptome profiling of the human cortex by MERFISH. (**A**) Schematic of 4000-gene MERFISH measurements of the human MTG and STG using a 48-bit error-correcting code. DAPI, 4′,6-diamidino-2-phenylindole; polyA, polyadenylate; HW, Hamming weight; HD, Hamming distance. (**B**) Example MERFISH images. (Left) MERFISH image of a single field of view, with maximum projection across all 48 bits. (Middle) Zoomed-in view of the boxed region in the left panel. (Right) Decoded RNA molecules of the region shown in the middle panel. Scale bars indicate the real size of the sample before expansion. (**C**) Cell type classification of the MTG and STG from MERFISH data and the expression of a subset of marker genes. EXC, excitatory neurons; INH, inhibitory neurons; ASC, astrocytes; MGC, microglial cells; OGC, oligodendrocytes; OPC, oligodendrocyte progenitor cells; ENDO, endothelial cells; MURAL, mural cells; IT, intratelencephalic-projecting neurons; ET, extratelencephalic-projecting neurons; NP, near-projecting neurons; CT, cortico-thalamic projecting neurons. The size and color of each dot correspond to the percentage of cells expressing the gene in each cluster and the average normalized expression level, respectively. (**D**) Proportions of excitatory neurons, inhibitory neurons, and major subclasses of non-neuronal cells in human MTG and STG and four mouse cortical regions (MOp, VIS, AUD, and TEa). (**E**) Proportion of subclasses of excitatory neurons (left), IT neurons (middle), and inhibitory neurons (right) in human MTG and STG and the four mouse cortical regions.

Cell compositions of the human and mouse cortices

Quantitative analysis of the cell composition using the MERFISH data showed that the human MTG and STG (white matter excluded) were composed of 26% excitatory neurons, 11% inhibitory neurons, and 63% non-neuronal cells (Fig. 1D). The excitatory neurons were predominantly intratelencephalic (IT) neurons (~93%), with only a small fraction of non-IT neurons (L6 CT, L5 ET, L5/6 NP, and L6b cells) (Fig. 1E, left). The IT neurons were subdivided into 46% L2/3 IT, 18% L4/5 IT, 19% L5 IT, 13% L6 IT, and 4% L6 IT *CAR3* cells (Fig. 1E, middle). The inhibitory neurons were composed of 13% *LAMP5*, 26% *PVALB*, 30% *SST*, and 31% *VIP* cells (Fig. 1E, right).

Next, we compared cell composition in human and mouse cortices. The human STG contains the auditory cortex, whereas the human MTG does not have a counterpart in mice, with the mouse temporal association area (TEa) considered to be the closest ortholog. We thus considered two MERFISH datasets that cover several regions of the mouse cortex: (i) our recently reported 258-gene MERFISH dataset of the primary motor cortex (MOp) (*17*) and (ii) a dataset generated from additional MERFISH experiments on a more posterior part of the mouse cortex, containing the visual cortex (VIS), auditory cortex (AUD), and TEa (fig. S8) using a similar gene panel and experimental protocol as for the MOp (*17*). Similar cell compositions were observed across these different mouse cortical regions (Fig. 1, D and E).

However, the cell composition of the human MTG and STG was substantially different from that of the mouse cortical regions. We observed a lower proportion of excitatory neurons and a higher proportion of glial cells (including astrocytes, oligodendrocytes, OPCs, and microglia) in the human cortex (Fig. 1D). The glia-to-neuron ratio was 1.4, consistent with results from various human cortical regions determined by other cell-counting methods (*18*, *19*) and five

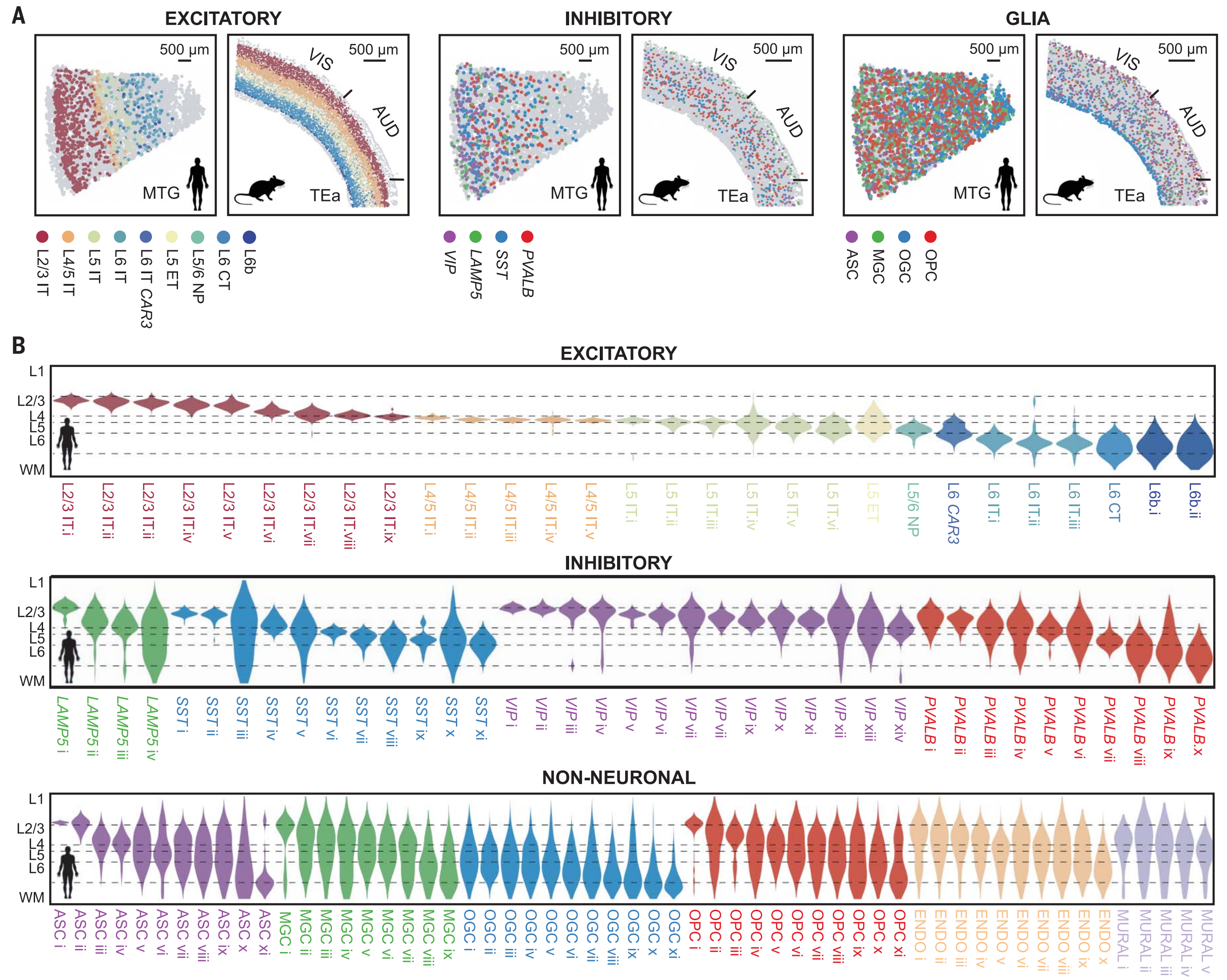


Fig. 2. Laminar organization of cell types in the human and mouse cortices. (**A**) Spatial maps of subclasses of excitatory neurons, inhibitory neurons, and glial cells determined by MERFISH in a human MTG slice and a mouse slice containing VIS, AUD, and TEa. Indicated subclasses are shown in colors; other cells are in gray. (**B**) Cortical-depth distribution of excitatory (top), inhibitory (middle), and non-neuronal (bottom) clusters in the human MTG. The dashed lines mark the approximate layer boundaries. WM, white matter.

times the ratio observed in mice (Fig. 1D) (*17*, *20*). The excitatory-to-inhibitory neuron ratio was 2.3 in humans, in line with recent independent measurements (*9*, *11*) and one-third of the ratio observed in mice (Fig. 1D) (*17*, *20*).

Among the excitatory neurons, the non-IT neuron proportion dropped from 29% in mice to 7% in humans (Fig. 1E, left), consistent with recent observations that L5 ET and L6 CT are less abundant in primates than in mice (*11*). The dominance of IT neurons in humans suggests an increased emphasis of intracortical communications. For inhibitory neurons, we observed a decrease in the proportion of *PVALB* neurons and an increase in the proportion of *VIP* neurons in humans relative to mice (Fig. 1E, right). In behaving animals, *VIP* interneurons regulate inhibition of excitatory neurons through inhibition of other interneurons, and such disinhibition facilitates modulation of sensory response and network dynamics by behavioral state and learning (*21*). The observed increase in *VIP* interneuron proportion thus suggests a potential mechanism for the enhanced capability of state-dependent sensory processing and learning-related neuronal dynamics in humans.

Spatial organizations of cells in the human and mouse cortices

In situ identification of cell types by MERFISH allowed us to map their spatial organizations. In humans, we observed a laminar organization of IT neurons across the cortical depth, whereas other excitatory neurons (including L5 ET, L5/6 NP, L6 CT, and L6b) were populated mostly in the deep layers (Fig. 2 and figs. S9 and S10), as expected (*8*, *9*, *22*). Among inhibitory neurons, *VIP* and *LAMP5* were enriched in the upper layers (L1 to L3), whereas *PVALB* and *SST* were more broadly distributed across the layers (Fig. 2 and figs. S9 and S10), consistent with previous observations (*8*, *9*, *22*). At the cluster level, inhibitory neurons also adopted a laminar organization, with many inhibitory clusters primarily restricted to one cortical layer or even a subportion of a layer (Fig. 2B, middle, and fig. S9), enriching and refining the knowledge of layer-restricted inhibitory neuron distributions (*8*). These spatial organizations of neurons were largely similar to those observed in the mouse cortex (figs. S10 and S11) (*17*).

Despite the overall conservation of laminar organization, we also found differences between humans and mice for some neuronal cell types. For instance, the L6b neurons were broadly dispersed in L6 and extended into L5 and white matter in the human MTG and STG, whereas in mice L6b formed a thin layer at the bottom of L6 (Fig. 3A), consistent with

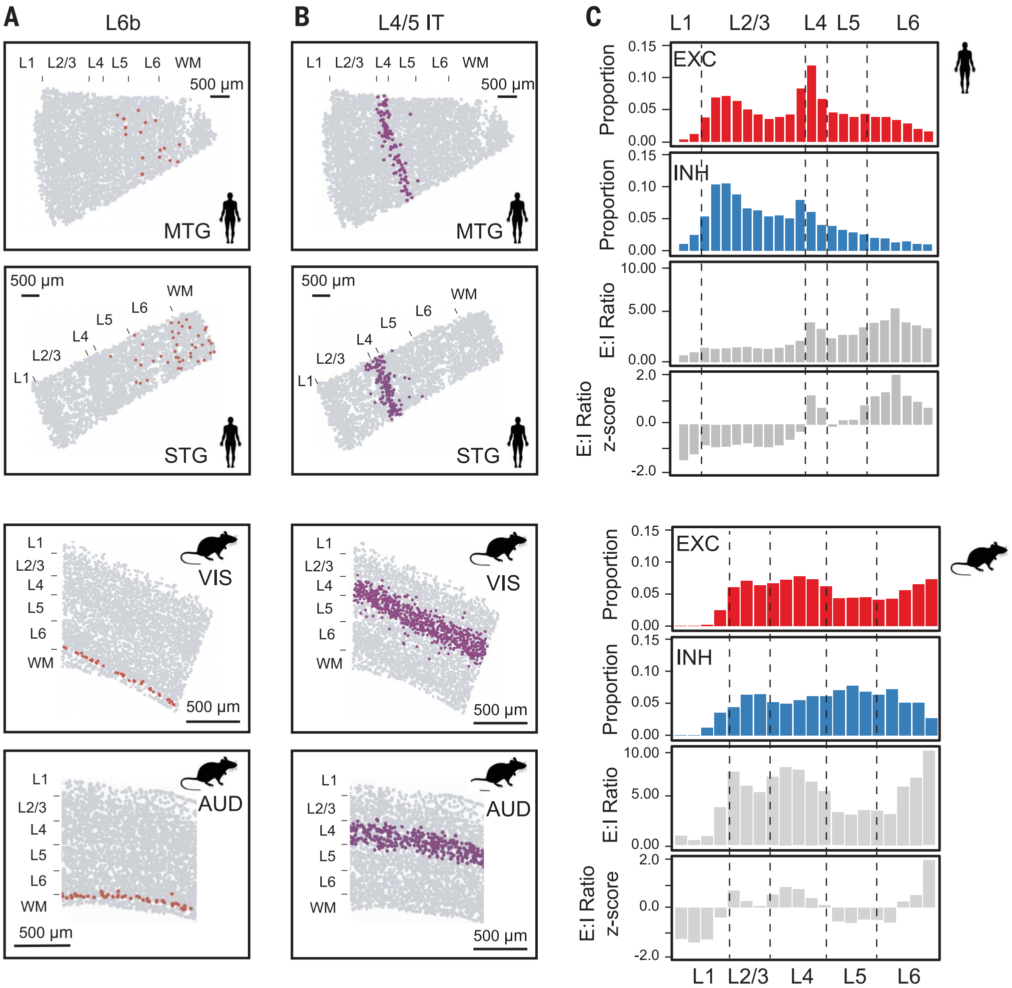


Fig. 3. Cortical-depth distributions of L6b, L4/5 IT, and excitatory-to-inhibitory neuronal ratio. (**A** and **B**) Spatial maps of L6b (A) and L4/5 IT (B) neurons in a human MTG slice (top), a human STG slice (upper middle), a VIS-containing region in a mouse slice (lower middle), and an AUD-containing region in a mouse slice (bottom). (**C**) Normalized cortical-depth distributions of excitatory (EXC) and inhibitory (INH) neurons, E:I ratio (i.e., the ratio of excitatory to inhibitory neurons), and E:I ratio *z* score in human (top) and mouse (bottom) cortices.

previous findings (*22*, *23*). The L4/5 IT neurons formed a dense and thin layer in the human MTG and STG, giving rise to a substantially higher density of excitatory neurons in L4 (Fig. 3, B and C, top). By contrast, the density of excitatory neurons in the mouse cortex was more uniform across L2/3 to L6 (Fig. 3C). L4 is known to vary between different cortical regions, so whether this difference is region- or species-specific remains an open question, although the several mouse cortical regions that we examined exhibited a similar density profile. We also observed a different cortical-depth dependence for the excitatory-to-inhibitory neuron ratio between humans and mice (Fig. 3C).

Non-neuronal cells also exhibited laminar organization in the human cortex. Oligodendrocytes were enriched in the deeper layers and white matter and depleted in the upper layers (L1 to L3) (fig. S10) (*9*). Although astrocytes, microglia, OPCs, endothelial cells, and mural cells were dispersed across all cortical layers at the subclass level (fig. S10), these cell types exhibited laminar organization at the cluster level (Fig. 2B, bottom). For example, the ASC i cluster was localized in L1, likely representing interlaminar astrocytes (*24*); ASC ii was enriched in L1 and L2/3; ASC iii and ASC iv were enriched in L2/3 and L4; ASC v to ix were dispersed across L2 to L6; and ASC x and xi were enriched in L6 and white matter (Fig. 2B, bottom), improving our understanding of astrocyte diversity and organization (*8*, *25*, *26*). Similarly, nearly all non-neuronal subclasses exhibited a gradually evolving cell composition across the cortical depth (Fig. 2B, bottom).

Cell-cell interactions in the human and mouse cortices

High-resolution MERFISH measurements of the spatial relationship between cells allowed us to predict cell-cell interactions arising from somatic contact or paracrine signaling (*27*, *28*), which can be inferred from soma contact or proximity that occurred at a higher frequency than random chance. Our MERFISH images showed frequent somatic contact or proximity between cells (Fig. 4A and fig. S12, A and B). Although the cell density in the human cortex was one-third of that in mice (fig. S12C), the median centroid distance between nearest-neighbor cells in humans was nearly identical to that in mice and comparable to the mean soma size (fig. S12, B and D), suggesting that specific mechanisms may exist to maintain or enhance cell-cell interactions in the expanded human cortex. To examine whether these potential cell-cell interactions were cell type specific, we considered cell types at the subclass level and calculated the frequency at which soma contact or proximity, determined on the basis of centroid distance (fig. S12A), was observed between two subclasses of cells. We then determined whether this frequency was significantly greater than random chance, thus reflecting an enrichment, by comparing the observed frequency with the expected frequencies from random spatial permutations that disrupted the spatial relationship between neighboring cells while preserving the local density of each cell type (figs. S13 to S15).

We observed cell type–specific patterns for soma contact or proximity enrichment in the human cortex that were different from those in the mouse cortex (Fig. 4B and fig. S16). Similar human-mouse differences were observed when we used segmented cell boundaries instead of distances between cell centroids to determine contacting cell pairs (fig. S17).

Inhibitory neurons and some deep-layer excitatory neurons in humans showed a tendency

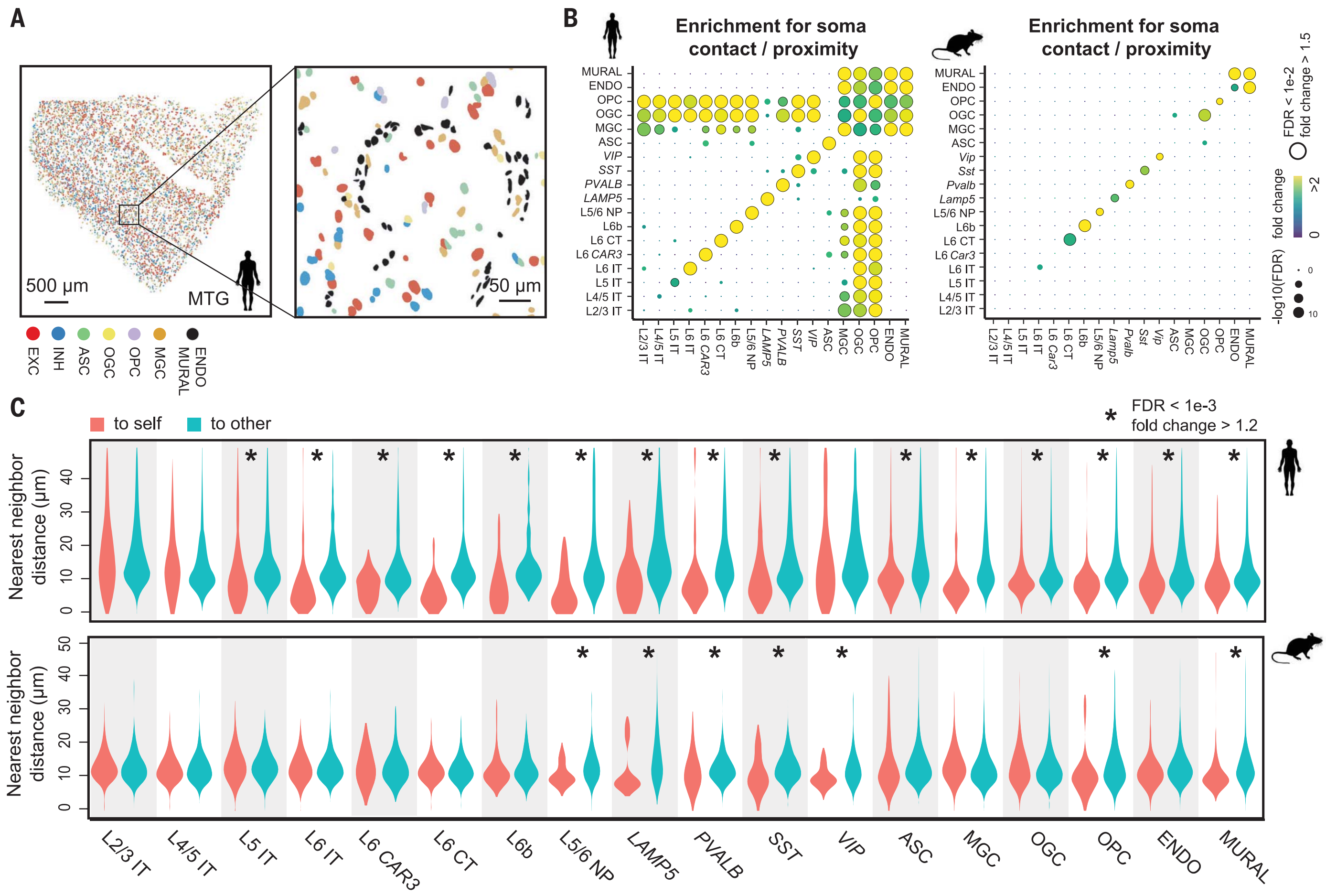


Fig. 4. Cell type–specific cell-cell interactions in the human and mouse cortices. (**A**) Spatial map of excitatory neurons, inhibitory neurons, and six major subclasses of non-neuronal cells in a human MTG slice (left) and a zoomed-in view of the boxed region (right). Colored shapes are cell nuclei segmentations. (**B**) Enrichment map of pairwise soma contact or proximity for subclasses of cells in human (left) and mouse (right) cortices. Dot color indicates the fold change between the observed frequency of soma contact or proximity and the average expected frequency from the spatial permutations that disrupt the spatial relationship between neighboring cells (fig. S13). Dot size indicates the significance level of the enrichment. False discovery rate (FDR): *P* value determined with upper-tailed *z* test and adjusted to FDR by the Benjamini-Hochberg (BH) procedure. (**C**) Distributions of the nearest-neighbor distances from cells in individual subclasses to cells in the same subclass ("to self"; red) or other subclasses ("to other"; blue) in human (top) and mouse (bottom) cortices. FDR: *P* value determined with the Wilcoxon rank-sum one-sided test and adjusted to FDR by the BH procedure.

to form contacting or proximity pairs among cells within the same subclass (Fig. 4B, left, and fig. S17). These results were further supported by examining the distances from individual neurons to their nearest neighbors in the same or different types (Fig. 4C, top). This tendency was also observed in mice but to a lesser degree for some neuronal types (Fig. 4, B and C), consistent with the previous observation that inhibitory neurons in mice tend to form intrasubtype nearest-neighbor pairs (*29*). Some non-neuronal cell types also exhibited such tendency for intratype soma proximity, but with noticeable differences between humans and mice. For example, we observed enrichment for soma contact or proximity among astrocytes in humans but not in mice (Fig. 4B and fig. S17). It has been shown that the processes of neighboring astrocytes intermingle substantially more in humans than in mice (*25*, *26*, *30*). Whether these observations are related to our findings here remains an open question.

A notable difference between humans and mice was observed for glial-vascular interactions. The human, but not mouse, cortex exhibited enrichment for soma contact or proximity between glial and vascular cells (Fig. 4B and fig. S17). MERFISH images showed that the cell bodies of oligodendrocytes and microglia were often clustered around vascular structures formed by endothelial and mural cells (Fig. 5A). These observations are corroborated by a recent electron microscopy study (*30*), which showed that oligodendrocyte and microglial cell bodies are adjacent to blood vessels, whereas astrocytes contact blood vessels primarily with their end feet but not cell bodies. Quantifications of MERFISH images showed that more microglia and oligodendrocytes, but not astrocytes, formed somatic contacts with blood vessels in humans than in mice (Fig. 5B and fig. S18).

Cross-species differences in cell-cell interactions were also observed between neurons and glial cells—in particular, oligodendrocytes and microglia. We observed substantial enrichment for soma contact or proximity between neurons and oligodendrocytes, including both mature oligodendrocytes and OPCs, in humans (Fig. 4B and fig. S17). Although somatic contacts between neurons and oligodendrocytes were also observed in mice and could represent bona fide interactions, the frequency of such events did not significantly exceed that expected from random chance. Moreover, a single neuron often formed contacts with several oligodendrocytes and OPCs in humans, whereas such

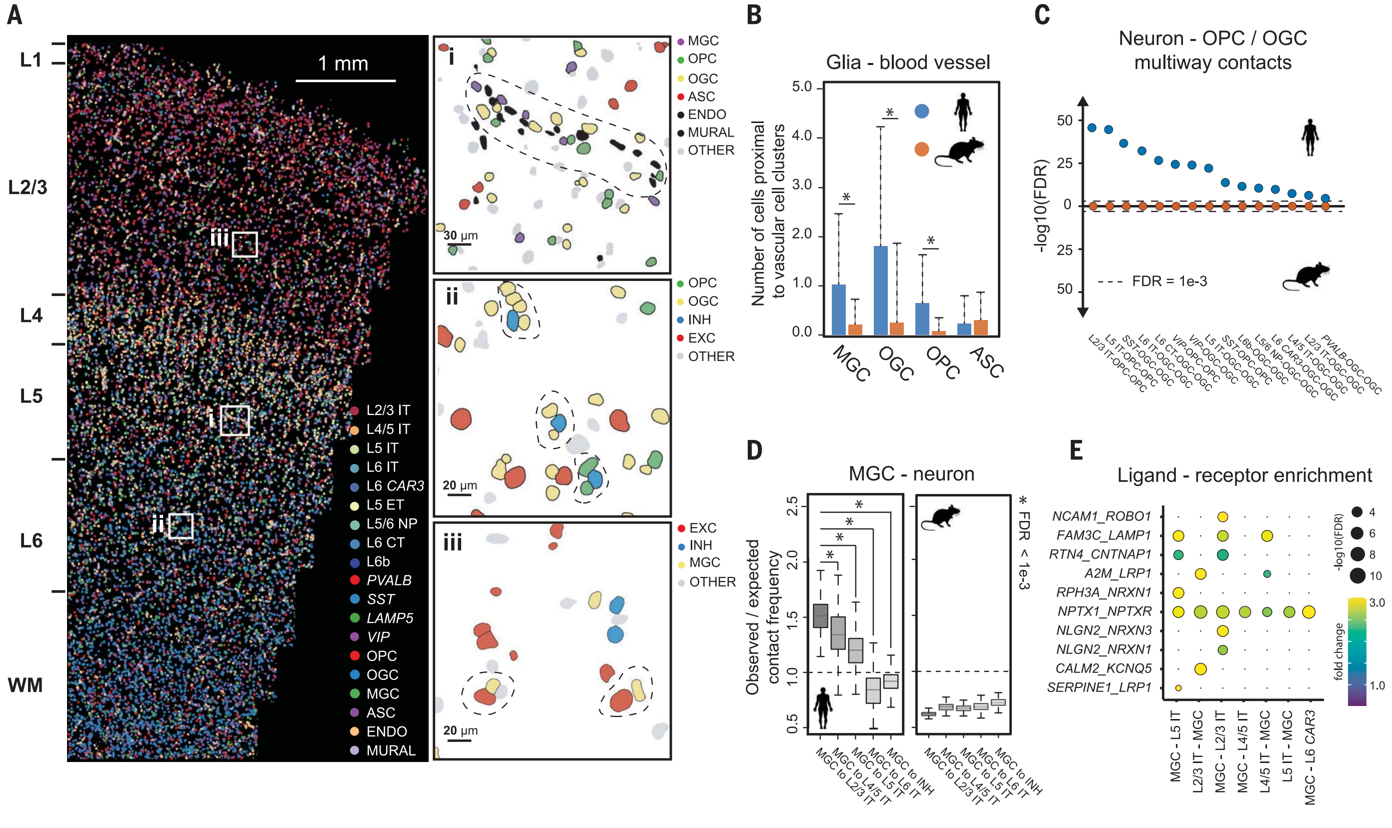


Fig. 5. Interactions between glial and vascular cells and between glial cells and neurons in the human and mouse cortices. (**A**) Spatial map of subclasses of cells in a human STG slice. (Top right) Zoomed-in view of boxed region i. A blood vessel with juxtaposed glial cells is indicated by a dashed line. (Middle right) Zoomed-in view of boxed region ii. Multiway contacts between neurons and oligodendrocytes and/or OPCs are indicated by dashed lines. (Bottom right) Zoomed-in view of boxed region iii. Contacting pairs of neurons and microglia are indicated by dashed lines. Colored and gray shapes are cell nuclei segmentations. (**B**) Average numbers of microglia, oligodendrocytes, OPCs, and astrocytes adjacent to each identified blood vessel in humans (blue) and mice (orange). Error bars denote SDs (n = 3415 vascular structures). *FDR < 1 × 10^{-3} (as determined in Fig. 4B). (**C**) Significance level of multiway contacts between neurons and oligodendrocytes and/or OPCs in human (blue) and mouse (orange) cortices. The significance level was determined by comparing the observed contact frequency with the expected frequencies from spatial permutations as described in fig. S13. FDR: P values determined with an upper-tailed Z test and adjusted to FDR by the BH procedure. (**D**) Ratio between observed contact frequency and expected contact frequency (from spatial permutations) between microglia and L2/3 IT, L4/5 IT, L5 IT, L6 IT, and inhibitory neurons in humans (left) and mice (right). In each box plot, the midline is the median, box edges are the 75th and 25th percentiles, and whiskers extend 1.5 times the interquartile range. *FDR < 1 × 10^{-3} (as determined in Fig. 4B). (**E**) Enrichment of ligand-receptor pairs in contacting microglia and IT neurons. The color and size of the dots respectively correspond to the fold change and significance level of the observed ligand-receptor scores over their expected values. FDR was determined as indicated for (C).

multiway contacts were not enriched above random chance in mice (Fig. 5, A and C).

In humans, among OPCs, a specific subpopulation exhibited a higher tendency to contact neurons. Our analyses of both MERFISH and SMART-seq data (*8*) showed that ~50% of the OPCs expressed glutamate decarboxylase 1 (*GAD1*), a gene encoding an enzyme that synthesizes γ-aminobutyric acid (GABA), whereas glutamate decarboxylase 2 (*GAD2*) and the GABA transporter gene *VGAT* (*SLC32A1*) were not expressed in OPCs (fig. S19A). Compared with *GAD1*-negative OPCs, *GAD1*-positive OPCs contacted neurons at a higher frequency (fig. S19, B and C).

Finally, our data revealed differences in microglia-neuron interactions in humans and mice. In the human MTG and STG, microglia were frequently juxtaposed with neurons (Fig. 5A), likely representing satellite microglia (*31*). In addition, these satellite microglia exhibited a greater degree of enrichment for soma contact with or proximity to excitatory neurons as opposed to inhibitory neurons (Fig. 4B, left; fig. S17; and Fig. 5, A and D). Moreover, among excitatory IT neurons, the tendency to contact microglia decreased with cortical depth (Fig. 5D). By contrast, no significant enrichment in microglia-neuron contact was observed in the mouse cortex (Fig. 4B, fig. S17, and Fig. 5D).

Furthermore, we identified ligand-receptor pairs enriched in contacting pairs of microglia and IT neurons from the MERFISH data (Fig. 5E), further validated by single-molecule FISH measurements (fig. S20). Among these, several ligands and receptors are genetically associated with neurodegenerative diseases (Fig. 5E)—for instance, alpha-2-macroglobulin (*A2M*) is genetically associated with Alzheimer's disease (*32*), low-density lipoprotein receptor-related protein 1 (*LRP1*) is a master regulator of tau uptake and spread (*33*), and neurexin (NRXN1/3) is implicated in autism (*34*).

Discussion

In this study, we demonstrated 4000-gene MERFISH imaging of human brain tissues. Our MERFISH images enabled in situ identification of >100 neuronal and non-neuronal cell populations and comprehensive mapping of the spatial organization of these cells in the human MTG and STG, resulting in a molecularly defined and spatially resolved cell atlas with high granularity. The cell composition in these human cortical regions differed markedly from that observed in several mouse cortical regions. The spatial organization of cells showed both common and divergent features between humans and mice.

Although we cannot exclude the possibility that some of these differences are due to different cortical regions, we consider this less likely because the different mouse cortical regions that we assessed exhibited similar cell type compositions and organizations, and the same was true for the human MTG and STG.

These high–spatial resolution cell atlases allowed us to systematically characterize proximity-based somatic interactions in a cell type–specific manner and revealed differences in cell-cell interactions between humans and mice. The differences were particularly pronounced for interactions between neuronal and non-neuronal cells. We observed substantially increased enrichment for soma contact or proximity between neurons and oligodendrocytes in the human cortex compared with the mouse cortex. Perineuronal oligodendrocytes (*35*) can provide metabolic support to neurons (*36*). Hence, the observed increase in contact enrichment between oligodendrocytes and neurons may be a result of evolutionary adaptation to higher energy demands during the firing of individual neurons in the human brain (*37*). In addition, we observed preferential enrichment for contact or proximity between microglia and excitatory neurons, compared with inhibitory neurons, in the human cortex, whereas the mouse cortex did not exhibit significant enrichment for such microglia-neuron contact. Satellite microglia can help maintain tissue homeostasis (*38*), and human genetics evidence suggests that microglia play a protective role that lowers the incidence of some neurodegenerative diseases (*39*). Our observation may thus represent a functional interaction between microglia and excitatory neurons in humans. Some ligand-receptor pairs genetically associated with neurodegenerative diseases were enriched in contacting microglia-neuron pairs as compared with noninteracting microglia and neurons, suggesting a possible molecular basis underlying the observed microglia-neuron interactions and a potential connection of these cell-cell interactions to neurodegenerative diseases. Evolution of non-neuronal cells has been suggested to follow a more complex pattern than simply increasing the cell abundance; it also involves the diversification of glial cells (*40*). Our observations of the enhanced enrichment for interactions between neurons and glia in the human cortex further expand on this view.

REFERENCES AND NOTES

1. S. Herculano-Houzel, *Front. Hum. Neurosci.* **3**, 31 (2009).
2. J. H. Lui, D. V. Hansen, A. R. Kriegstein, *Cell* **146**, 18–36 (2011).
3. R. D. Fields, B. Stevens-Graham, *Science* **298**, 556–562 (2002).
4. B. A. Barres, *Neuron* **60**, 430–440 (2008).
5. C. Scuderi, A. Verkhratsky, in *Progress in Molecular Biology and Translational Science* (Elsevier, 2020), vol. 173, pp. 301–330.
6. H.-G. Bernstein, J. Steiner, B. Bogerts, *Expert Rev. Neurother.* **9**, 1059–1071 (2009).
7. M. W. Salter, B. Stevens, *Nat. Med.* **23**, 1018–1027 (2017).
8. R. D. Hodge *et al*., *Nature* **573**, 61–68 (2019).
9. C. M. Langseth *et al*., *Commun. Biol.* **4**, 998 (2021).
10. C. Luo *et al*., *Science* **357**, 600–604 (2017).
11. T. E. Bakken *et al*., *Nature* **598**, 111–119 (2021).
12. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, *Science* **348**, aaa6090 (2015).
13. Y. Sun, A. Chakrabartty, *Biochem. Cell Biol.* **94**, 545–550 (2016).
14. F. Chen, P. W. Tillberg, E. S. Boyden, *Science* **347**, 543–548 (2015).
15. C. Xia, J. Fan, G. Emanuel, J. Hao, X. Zhuang, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19490–19499 (2019).
16. G. Wang, J. R. Moffitt, X. Zhuang, *Sci. Rep.* **8**, 4847 (2018).
17. M. Zhang *et al*., *Nature* **598**, 137–143 (2021).
18. C. C. Sherwood *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13606–13611 (2006).
19. C. S. von Bartheld, J. Bahney, S. Herculano-Houzel, *J. Comp. Neurol.* **524**, 3865–3895 (2016).
20. D. Keller, C. Erö, H. Markram, *Front. Neuroanat.* **12**, 83 (2018).
21. J. J. Letzkus, S. B. E. Wolff, A. Lüthi, *Neuron* **88**, 264–276 (2015).
22. H. Zeng *et al*., *Cell* **149**, 483–496 (2012).
23. M. Judaš, G. Sedmak, M. Pletikos, *J. Anat.* **217**, 344–367 (2010).
24. J. A. Colombo, H. D. Reisin, *Brain Res.* **1006**, 126–131 (2004).
25. N. A. Oberheim *et al*., *J. Neurosci.* **29**, 3276–3287 (2009).
26. A. A. Sosunov *et al*., *J. Neurosci.* **34**, 2285–2298 (2014).
27. E. Armingol, A. Officer, O. Harismendy, N. E. Lewis, *Nat. Rev. Genet.* **22**, 71–88 (2021).
28. J. Fan, K. Slowikowski, F. Zhang, *Exp. Mol. Med.* **52**, 1452–1465 (2020).
29. X. Wang *et al*., *Science* **361**, eaat5691 (2018).
30. A. Shapson-Coe *et al*., A connectomic study of a petascale fragment of human cerebral cortex. bioRxiv 2021.05.29.446289 [Preprint] (2021); https://doi.org/10.1101/2021.05.29.446289.
31. V. Stratoulias, J. L. Venero, M. È. Tremblay, B. Joseph, *EMBO J.* **38**, e101997 (2019).
32. D. Blacker *et al*., *Nat. Genet.* **19**, 357–360 (1998).
33. J. N. Rauch *et al*., *Nature* **580**, 381–385 (2020).
34. T. C. Südhof, *Cell* **171**, 745–769 (2017).
35. N. Baumann, D. Pham-Dinh, *Physiol. Rev.* **81**, 871–927 (2001).
36. T. Philips, J. D. Rothstein, *J. Clin. Invest.* **127**, 3271–3280 (2017).
37. P. Lennie, *Curr. Biol.* **13**, 493–497 (2003).
38. Q. Li, B. A. Barres, *Nat. Rev. Immunol.* **18**, 225–242 (2018).
39. D. V. Hansen, J. E. Hanson, M. Sheng, *J. Cell Biol.* **217**, 459–472 (2018).
40. A. Verkhratsky, M. S. Ho, V. Parpura, in *Neuroglia in Neurodegenerative Diseases*, A. Verkhratsky, M. S. Ho, R. Zorec, V. Parpura, Eds. (Springer, 2019), vol. 1175 of *Advances in Experimental Medicine and Biology*, pp. 15–44.
41. R. Fang *et al*., Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH, Dryad (2022); https://doi.org/10.5061/dryad.x3ffbg7mw.
42. H. Babcock *et al*., ZhuangLab/storm-control: v2019.06.28 release, Zenodo (2019); https://doi.org/10.5281/zenodo.3264857
43. R. Fang, S. Eichhorn, G. Emanuel, H. Babcock, X. Zhuang, r3fang/MERlin: v0.2.0_20220413, Zenodo (2022); https://doi.org/10.5281/zenodo.6459546.

ACKNOWLEDGMENTS

We thank T. Bakken, W. Allen, and C. E. Ang for helpful comments on the manuscript and R. Hodge and J. Nyhus for assistance in postmortem sample preparation. **Funding:** This work is supported in part by the National Institutes of Health and the Chan Zuckerberg Initiative. R.F. is a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation. X.Z. is a Howard Hughes Medical Institute Investigator. **Author contributions:** E.S.L. and X.Z. conceived the study. R.F., C.X., and X.Z. designed experiments with input from J.L.C., M.Z., J.H., B.L., J.A.M., and E.S.L. M.Z. and J.H. performed human tissue MERFISH protocol testing. C.X. and J.A.M. designed the MERFISH gene panel. J.L.C. performed human tissue processing and quality testing. R.F., C.X., Z.H., and A.R.H. performed MERFISH experiments. R.F. and C.X. performed data analysis. R.F., C.X., J.L.C., B.L., J.A.M., E.S.L., and X.Z. evaluated experimental results. R.F. and X.Z. wrote the paper with input from C.X., J.L.C., M.Z., J.H., Z.H., A.R.H., B.L., J.A.M., and E.S.L. X.Z. oversaw the project. **Competing interests:** C.X. and X.Z. are inventors on patents applied for by Harvard University related to MERFISH. X.Z. is a cofounder and consultant of Vizgen. **Data and materials availability:** MERFISH data are available at Dryad (*41*). The SMART-seq data are available at https://portal.brain-map.org/atlases-and-data/rnaseq/human-mtg-smart-seq. All other data are in the main paper or supplementary materials. The MERFISH image acquisition software is available at Zenodo (*42*). The analysis software is available at Zenodo (*43*).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by the Howard Hughes Medical Institute, a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY 4.0 public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm1741
Materials and Methods
Figs. S1 to S20
Tables S1 to S6
References (*44*–*54*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 31 August 2021; accepted 11 May 2022
10.1126/science.abm1741

NEUROSCIENCE

A specific circuit in the midbrain detects stress and induces restorative sleep

Xiao Yu[1]*†‡§, Guangchao Zhao[2]†, Dan Wang[2], Sa Wang[2], Rui Li[2], Ao Li[2], Huan Wang[3,4], Mathieu Nollet[1,5], You Young Chun[1], Tianyuan Zhao[1], Raquel Yustos[1], Huiming Li[2], Jianshuai Zhao[2], Jiannan Li[2], Min Cai[6], Alexei L. Vyssotski[7], Yulong Li[3,4], Hailong Dong[2]*, Nicholas P. Franks[1,5]*, William Wisden[1,5]*

In mice, social defeat stress (SDS), an ethological model for psychosocial stress, induces sleep. Such sleep could enable resilience, but how stress promotes sleep is unclear. Activity-dependent tagging revealed a subset of ventral tegmental area γ-aminobutyric acid (GABA)–somatostatin (VTA$^{Vgat\text{-}Sst}$) cells that sense stress and drive non–rapid eye movement (NREM) and REM sleep through the lateral hypothalamus and also inhibit corticotropin-releasing factor (CRF) release in the paraventricular hypothalamus. Transient stress enhances the activity of VTA$^{Vgat\text{-}Sst}$ cells for several hours, allowing them to exert their sleep effects persistently. Lesioning of VTA$^{Vgat\text{-}Sst}$ cells abolished SDS-induced sleep; without it, anxiety and corticosterone concentrations remained increased after stress. Thus, a specific circuit allows animals to restore mental and body functions by sleeping, potentially providing a refined route for treating anxiety disorders.

Acute stress activates the hypothalamic–pituitary–adrenal axis, and the resulting fast increase in blood glucocorticoid concentrations aids immediate survival (*1*–*3*). But chronically increased concentrations of glucocorticoids are harmful (*1*, *2*), as can be memories of stressful experiences (*4*). Although stress can cause insomnia and raise stress hormones (*3*, *5*–*8*), the opposite is also true: Chronic stress increases rapid eye movement (REM) sleep (*9*); and sleep in rodents is induced by specific types of stress, such as social defeat stress (SDS). Although the function and benefits of sleep remain unclear, sleep is certainly restorative (*10*). Thus, sleep has been suggested to be one of the mechanisms for alleviating the malign effects of stress (*4*, *9*, *11*, *12*). Whether a specific circuit links stress and sleep is, however, unknown. We reasoned that the ventral tegmental area (VTA) in the midbrain could provide a link.

The VTA regulates reward, aversion, goal-directed behaviors, and social contact (*13*–*15*). It also influences responses to stress and threats (*16*, *17*), and strongly affects sleep and wake: VTA^{Vglut2} and VTATH neurons promote wakefulness (*18*, *19*), whereas VTA$^{Vgat\ or\ Gad67}$ neurons induce sleep (*18*, *20*, *21*). Because some γ-aminobutyric acid (GABA) VTA neurons are activated by stressful and aversive stimuli (*16*, *22*–*24*), we hypothesized that this route allows stress to induce sleep.

[1]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK. [2]Department of Anesthesiology and Perioperative Medicine, Xijing Hospital, Fourth Military Medical University, Xi'an, China. [3]State Key Laboratory of Membrane Biology, Peking University School of Life Sciences, Beijing 100871, China. [4]PKU-IDG/McGovern Institute for Brain Research, Beijing 100871, China. [5]UK Dementia Research Institute, Imperial College London, London SW7 2AZ, UK. [6]Department of Psychiatry, Xijing Hospital, Fourth Military Medical University, Xi'an, China. [7]Institute of Neuroinformatics, University of Zürich/ETH Zürich, Zürich, Switzerland.
***Corresponding author. Email: xiao.1.yu@kcl.ac.uk (X.Y.); hldong6@hotmail.com (H.D.); n.franks@imperial.ac.uk (N.P.F.); w.wisden@imperial.ac.uk (W.W.)**
†These authors contributed equally to this work.
‡Present address: Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College, London SE5 9RT, UK.
§Present address: UK Dementia Research Institute, King's College, London SE5 9RT, UK.

Social defeat stress induces sleep

We first assessed the sleep-wake architecture of mice after they had experienced either SDS from aggressors with consecutive episodes of SDS for 1 hour, or a control procedure, in which the experimental mouse (intruder) was separated from the resident aggressor by a clear partition (fig. S1A). As a further control, instead of an aggressor mouse, we introduced a juvenile male mouse as the resident for 1 hour (fig. S1B). During this nonstress procedure, the experimental mice experienced social interaction. As a control for whether physical activity induces sleep, the mice experienced voluntary wheel running or forced treadmill running continuously for 1 hour (fig. S1, C and D); alternatively, as another control, the mice were placed in an unfamiliar environment with a novel object (fig. S1E). Corticosterone (CORT) concentrations in mice increased after the SDS sessions (Fig. 1A), but not after exposures to juvenile mice, physical exercise, or a novel environment or objects (fig. S2, A, C, E, and G). After SDS, NREM sleep latency was shortened, and both NREM and REM sleep were continuously elevated for 5 hours (Fig. 1, B and C, and fig. S2I), consistent with previous observations (*11*, *25*). Mice that experienced a nonstressful procedure, voluntary wheel running, forced treadmill running or that were deprived of 1 hour of sleep by placing them in a novel environment, however, did not have induced sleep above baseline (fig. S2, A to H and J to M), suggesting that social interaction or physical exercise did not induce sleep and SDS procedures did not cause a sleep rebound while the mice were awake.

Sleep relieves SDS-induced anxiety and CORT concentrations independently

We explored potential functions of sleep after SDS. For mice allowed sufficient sleep (home cage sleep) after SDS, anxiety-like behaviors caused by SDS were rapidly reduced, as seen in the elevated plus maze and open-field assays (Fig. 1, D to F). If mild sleep deprivation over 4 hours took place immediately after SDS, the mice remained in an anxious state (Fig. 1, D to F). For mice allowed sufficient home cage sleep after SDS, raised CORT concentrations returned to baseline over 60 min (Fig. 1G). If mild sleep deprivation occurred immediately after stress, however, CORT concentrations remained elevated (Fig. 1G). But pharmacologically reducing the CORT concentrations induced by SDS during sleep deprivation, by using a corticosterone synthesis inhibitor, metyrapone (fig. S3A), did not reduce anxiety after sleep deprivation (fig. S3, B and C).

Identification of neurons activated by stress

To identify the circuitry that induces restorative sleep, we mapped cFOS expression throughout the brain. After the SDS protocol (fig. S1A), cFOS was strongly increased in brain areas involved in stress responses (fig. S4, A and B), including the VTA (Fig. 1H and fig. S4). In the VTA, cells activated (cFOS-positive) by stress predominantly expressed the GABAergic marker *Vgat* (60%) or GABA (57%) (Fig. 1I and fig. S5, A and B), whereas relatively fewer cells expressed the glutamatergic marker *Vglut2* (20%) or the dopaminergic marker tyrosine hydroxylase (TH) (10%) (fig. S5, C and D). However, physical exercise did not induce cFOS in the VTA as a whole, and particularly not in VTA^{Vgat} cells, but forced treadmill running slightly increased cFOS expression in TH-positive cells (fig. S5, E to G). For the subsequent studies, we focused on the VTA^{Vgat} neurons, as only these induce sleep (*18*).

VTA^{Vgat} neurons have persistently increased activity in response to SDS

VTA^{Vgat} neurons rapidly and strongly responded when mice experienced an attack during SDS (Fig. 1, J and K), as assessed by GCaMP6 fiber photometry. The cells did not respond when the mice were presented with novel objects or placed in an unfamiliar environment (fig. S6, A and B). During SDS, the calcium signal in VTA^{Vgat} neurons increased and stayed enhanced for about 5 hours (Fig. 1L), correlating with the behavioral result of prolonged sleep after SDS (Fig. 1, B and C). By contrast, voluntary wheel

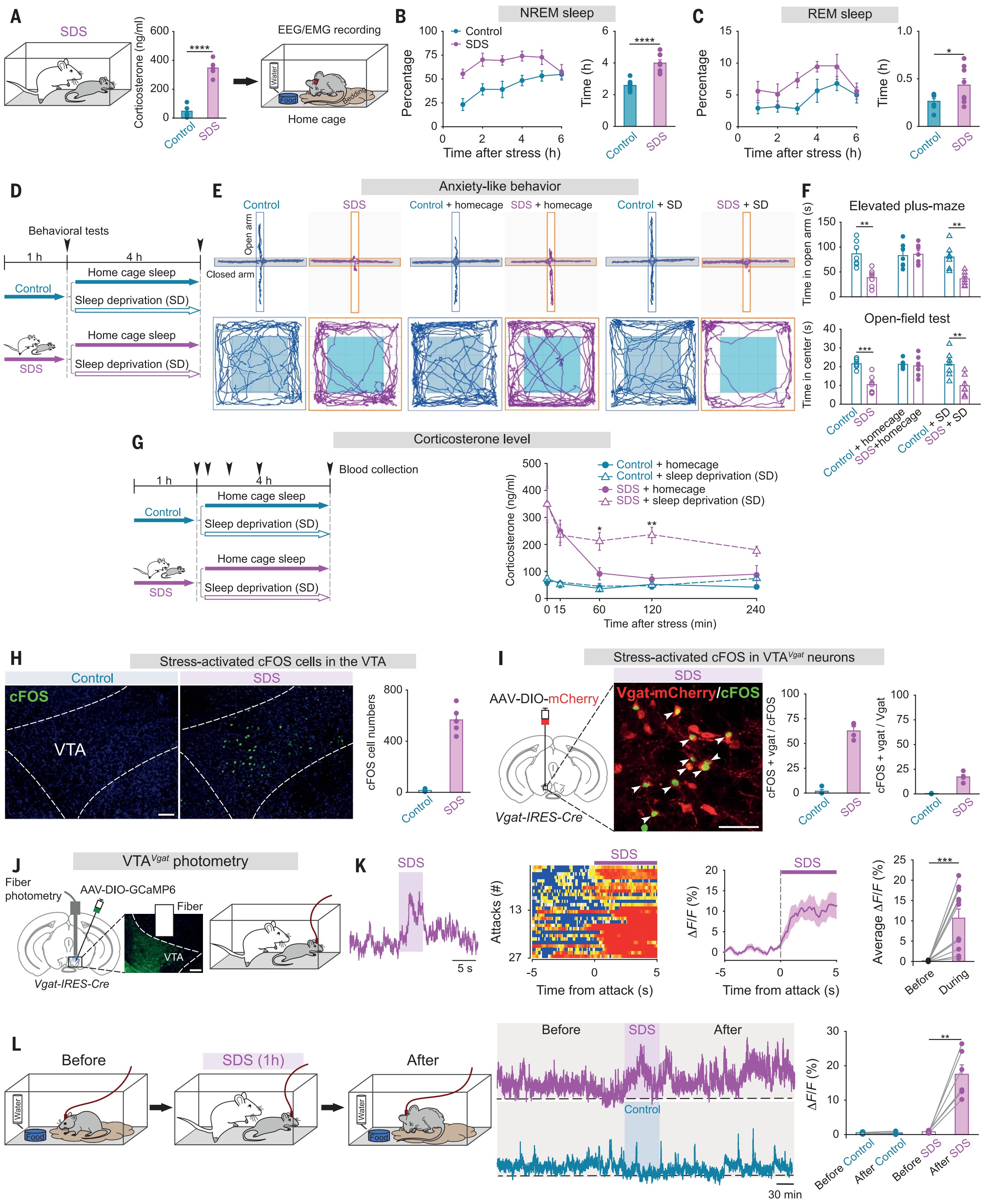


Figure 1 continued on next page

Fig. 1. Stress increases sleep whereas sleep reduces SDS-induced anxiety and stress activates VTA^{Vgat} neurons. (**A**) Experimental procedure and corticosterone concentrations (n = 6 mice per group). (**B** and **C**) Percentage and time of NREM (B) and REM (C) sleep after control or SDS (n = 8 mice per group). (**D** to **F**) Plan of the experimental procedure (D), tracing of locomotion for representative animals (E), time spent in the open arms of the elevated plus maze and in the center zone during the open-field test (F) (n = 7 mice per group). (**G**) Plan of the experimental procedure and corticosterone concentrations (n = 6 mice per group). (**H** and **I**) cFOS expression and quantification in the VTA after control or SDS (n = 5 mice per group) (H) or in genetically labeled VTA^{Vgat} neurons (n = 4 mice per group). (I). Arrowheads indicate double-labeled cells. Scale bar, 100 μm. (**J** and **K**) Fiber photometry setup and GCaMP6 expression in VTA^{Vgat} neurons (J). Fiber photometry measuring calcium signals responding to SDS (n = 11 mice, 27 trials). Raw calcium signal traces, color matrix of signals for all trials, $\Delta F/F$ ratios across the experimental period and average $\Delta F/F$ ratios before and during the procedure (K). Scale bar, 100 μm. (**L**) Fiber photometry measuring long-term calcium signals in VTA^{Vgat} neurons. Traces across the experimental procedure and average $\Delta F/F$ ratios before and after the procedures (n = 6 mice per group). (A to C) unpaired t test, *$p < 0.05$, ****$p < 0.0001$; (F and G) Two-way analysis of variance (ANOVA) with bonferroni post hoc test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; (K and L) Paired t test, **$p < 0.01$, ***$p < 0.001$.

running, forced treadmill running, or a novel environment did not affect baseline activity of VTA^{Vgat} neurons (fig. S6, C and D).

Subsets of VTA^{Vgat} neurons mediate SDS-induced sleep

Because only a subset of VTA^{Vgat} neurons (20%) were excited by SDS (Fig. 1I), we undertook cFOS-dependent activity tagging linked to expression of DREADD hM3Dq-mCherry to test whether this VTA subset could induce sleep (Fig. 2A). Mice experienced either SDS or a nonstressful procedure, voluntary wheel running or forced treadmill running, while the VTA^{Vgat} neurons were selectively activity-tagged with Cre-recombinase–dependent tagging vectors (*AAV-cFOS-tTA* and *AAV-TRE-DIO-hM3Dq-mCherry*; Fig. 2A and fig. S7, A, D, and G). Compared with pan-VTA^{Vgat} neurons expressing mCherry, only 15% of the VTA^{Vgat} neurons were captured by activity tagging during SDS (Fig. 2B). We then reactivated these SDS-tagged VTA^{Vgat} neurons with clozapine N-oxide (CNO). Chemogenetic reactivation decreased sleep latencies and increased sleep times (Fig. 2, C and D). Thus, reactivation of SDS-activated VTA^{Vgat} neurons recapitulated sleep architectures induced by SDS (Fig. 2E). Of note, a few cells (2.6%) were tagged during the nonstressed procedures (Fig. 2B). However, chemogenetic reactivation of these particular tagged VTA^{Vgat} cells did not elicit sleep (fig. S7, B and C). Moreover, only rare cells were tagged when mice experienced physical exercise (Fig. 2B), and therefore chemogenetic reactivation did not induce sleep (fig. S7, D to I).

To examine the necessity of VTA^{Vgat} subsets for SDS-induced sleep, we chemogenetically inhibited SDS-tagged VTA^{Vgat} neurons using cFOS-dependent expression of DREADD hM4Di-mCherry. *AAV-cFOS-tTA* and *AAV-TRE-DIO-hM4Di-mCherry* were injected into the VTA of *Vgat-IRES-Cre* mice (Fig. 2F). Mice were subjected to SDS (first stress episode) to allow VTA^{Vgat} neurons to become tagged with hM4Di-mCherry, then given CNO to inhibit the tagged neurons, and mice were subsequently challenged with a second bout of SDS (second stress episode), followed by measurement of their sleep profile (Fig. 2F). SDS-induced sleep was diminished after chemogenetically inhibiting tagged VTA^{Vgat} neurons (Fig. 2, G and H).

Circuits linking stress and sleep

We next investigated the circuitry linking SDS and VTA^{Vgat}-induced sleep. We expressed GCaMP6 selectively in VTA^{Vgat} cells and used fiber photometry to measure how the VTA^{Vgat} terminals in different locations responded to stress. Only the terminals of the VTA^{Vgat} cells projecting to the lateral hypothalamus (LH) had increased Ca^{2+} signals after SDS (Fig. 3, A and B), whereas the VTA^{Vgat} projections in the CeA, LHb, and hippocampal dentate granule (DG) cells showed no responses (fig. S8). To determine the function of the VTA^{Vgat}→LH pathway activated by stress on sleep, we injected *retro-AAV-TRE-DIO-Flpo* into the LH, together with the injection of *AAV-cFOS-tTA* and *AAV-fDIO-hM3Dq-mCherry* into the VTA of *Vgat-IRES-Cre* mice (Fig. 3C). Following intersectional activity tagging during SDS and chemogenetic reactivation, the VTA→LH pathway promoted sleep (Fig. 3, D and E). The hM3Dq-mCherry labeling produced in the VTA^{Vgat} neurons of this experiment mainly traced out axons to the LH (fig. S9).

We used optogenetics to confirm the above result. The behavioral experiments were repeated by using cFOS-based activity tagging with ChR2 delivered into the VTA of *Vgat-IRES-Cre* mice (fig. S10A). VTA^{Vgat} neurons became selectively ChR2-tagged during SDS (fig. S10A). Mapping of VTA^{Vgat} projections by injecting *AAV-DIO-ChR2-EYFP* into *Vgat-IRES-Cre* mice showed broad projections (fig. S10B) (*18*). However, those SDS-tagged VTA^{Vgat} cells detected with ChR2 activity mapping primarily innervated the LH (fig. S10C). When SDS-ChR2–tagged terminals in the LH of the VTA^{Vgat}→LH pathway were reactivated by optogenetic stimulation, this elicited NREM sleep from waking (fig. S10, D and E).

Stress-driven input-output organizations

We investigated the identity and activity of VTA^{Vgat} afferents relevant for stress using a rabies system, combined with activity mapping. VTA^{Vgat} neurons were seeded with rabies coat protein and its receptor by injecting *AAV-DIO-N2cG* and *AAV-DIO-TVA-nGFP*, followed by injection of *RABV-N2cΔG-EnvAmCherry* into the VTA (Fig. 3F). The animals were then given control experiences or SDS, respectively. Then we conducted brain-wide mapping of rabies-labeled presynaptic inputs and stress-activated cFOS expression (Fig. 3F). cFOS was induced by stress in many brain regions (fig. S4), and from the rabies tracing, VTA^{Vgat} inputs originated in many locations (fig. S11) (*26*). However, only the lateral preoptic (LPO), paraventricular hypothalamus (PVH), and periaqueductal grey (PAG) areas had overlap with cFOS-positive cells and rabies-labeled VTA^{Vgat} inputs (Fig. 3, G, H, and L, and figs. S12, A to C, and S13A).

We determined the inputs of VTA^{Vgat} neurons that project to the LH. *AVV-DIO-N2cG* and *AAV-DIO-TVA-nGFP* were seeded as before in VTA^{Vgat} neurons, and *RABV-N2cΔG-EnvAm-Cherry* was injected into the terminal fields of the VTA^{Vgat} neurons in the LH. As before, the mice were given control experiences or SDS. Then we mapped cFOS expression (Fig. 3I). We obtained an identical result as above: LH-projecting VTA^{Vgat} neurons received stress-activated inputs from the LPO, PVH, and PAG (Fig. 3, J to L, and figs. S12, D to F, and S13B).

We further determined if these stress-activated inputs were specific to the stress-activated VTA^{Vgat} subset. The SDS-activated VTA^{Vgat} cells were specifically ablated with Casp3 by using activity tagging (*AAV-cFOS-tTA* and *AAV-TRE-DIO-Casp3*). Then we conducted rabies tracing and activity mapping (fig. S14A). Ablation of SDS-activated VTA^{Vgat} cells largely reduced the stress-driven inputs (cFOS/rabies) (fig. S14, B and C, and Fig. 3L).

VTA somatostatin neurons are necessary for SDS-induced sleep

Given that GABAergic VTA cells are heterogeneous (*13*, *27*, *28*), and only a subset of VTA^{Vgat} cells responded to SDS (Fig. 2B), we looked for subtypes of VTA^{GABA} cells responsible for SDS-induced sleep. First, we examined by single-cell quantitative polymerase chain reaction the molecular identities of SDS-tagged

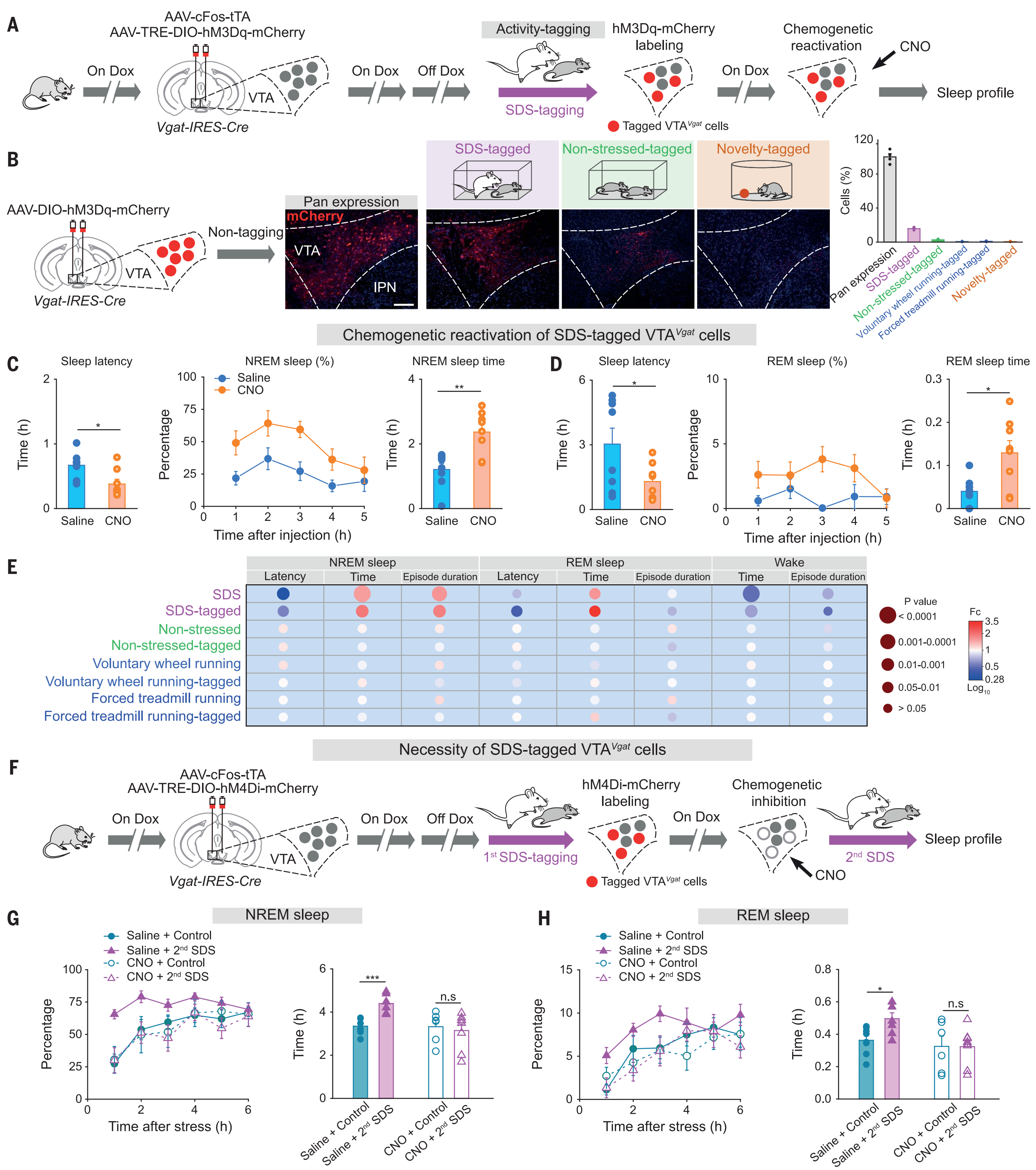


Fig. 2. Sufficiency and necessity of the stress-activated VTA^{Vgat} neurons for SDS-induced sleep. (**A**) Activity-tagging protocol for testing the sufficiency of SDS-activated VTA^{Vgat} cells for sleep. (**B**) Expression and quantification of pan or activity-tagged hM3Dq-mCherry transgene in VTA^{Vgat} neurons (n = 4 mice per group). Scale bar, 100 μm. (**C** and **D**) Chemogenetic reactivation of tagged VTA^{Vgat} neurons for sleep (n = 8 mice per group). Graphs show sleep latency, percentage and time of NREM (C) or REM (D) sleep. Unpaired t test, $*p < 0.05$, $**p < 0.01$. (**E**) Matrix bubble summary shows fold change (Fc) of sleep parameters after SDS, nonstressed, voluntary wheel running, forced treadmill running or the chemogenetic reactivation of tagged VTA^{Vgat} neurons. (**F**) Activity-tagging protocol for testing the necessity of SDS-activated VTA^{Vgat} cells for sleep. (**G** and **H**) Percentage and time of NREM (G) or REM sleep (H) in mice given second SDS after chemogenetic inhibition of first SDS-tagged VTA^{Vgat} cells (n = 6 to 8 mice per group). Two-way ANOVA with bonferroni post hoc test. $*p < 0.05$, $***p < 0.001$, n.s: not significant.

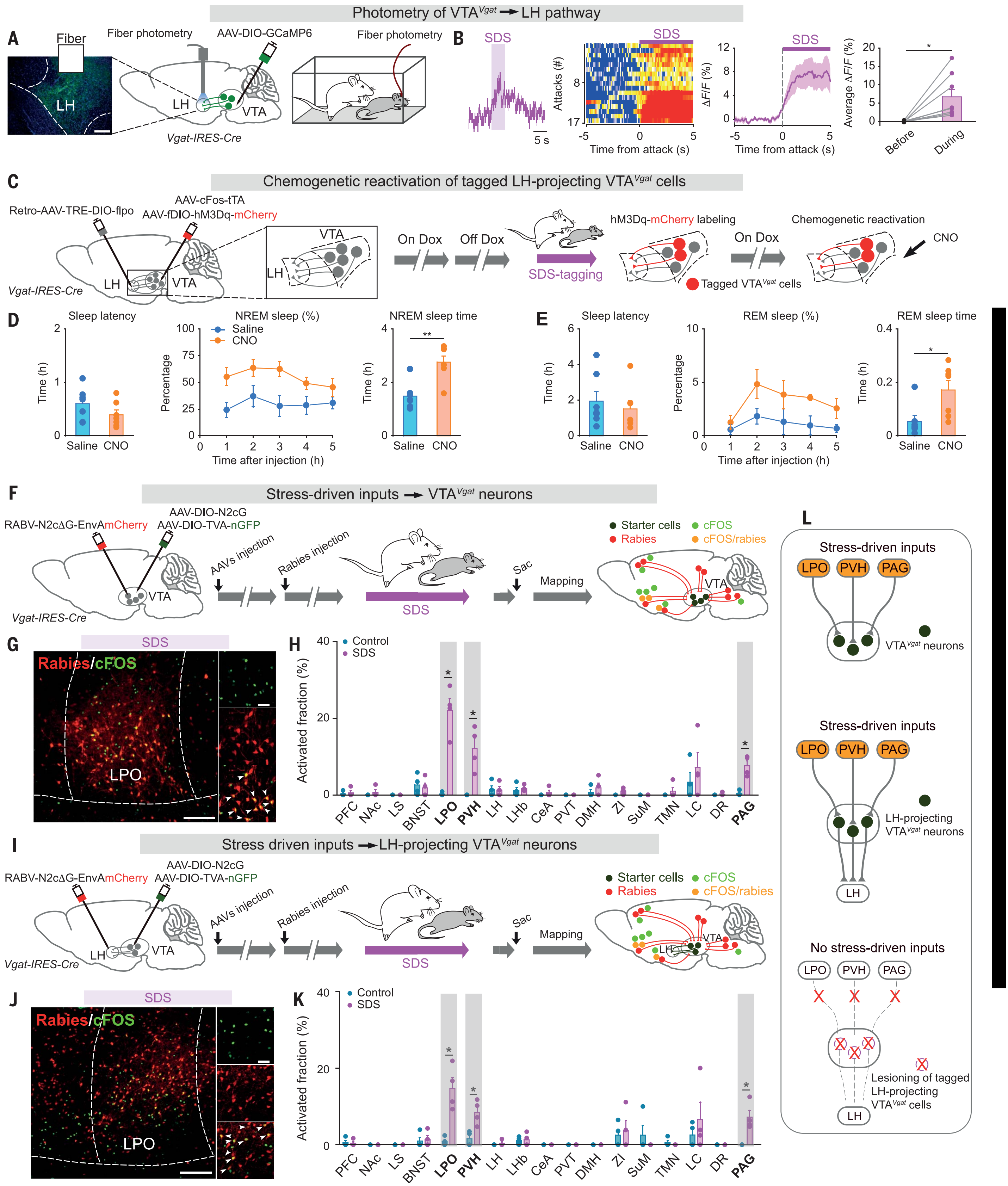


Figure 3 continued on next page

Fig. 3. Input-output circuitry linking stress and sleep. (**A** and **B**) Fiber photometry measuring terminal calcium signals of the VTA^{Vgat}→LH pathway responding to SDS (A). Raw traces, color matrix of GCaMP6 signals of VTA^{Vgat}→LH for all trials, $\Delta F/F$ ratios across the experimental period and average $\Delta F/F$ before and during SDS (n = 8 mice, 17 trials) (B). Paired t test, $*p < 0.05$. Scale bar, 100 μm. (**C**) Activity-tagging protocol for reactivating the SDS-tagged LH-projecting VTA^{Vgat} cells. (**D** and **E**) Sleep latency, percentage and time of NREM (D) and REM (E) sleep after reactivation of SDS-tagged LH-projecting VTA^{Vgat} cells (n = 7 mice per group). Unpaired t test, $*p < 0.05$, $**p < 0.01$. (**F**) Rabies virus–based retrograde tracing for identification of stress-driven inputs to VTA^{Vgat} neurons. (**G**) Immunostaining images showing presynaptic inputs to VTA^{Vgat} neurons from LPO and cFOS-positive cells activated by SDS. Scale bar, 200 μm and 50 μm (inset). (**H**) Summary statistics of activated fractions (cFOS/rabies double-labeled cells/total rabies-positive cells) (n = 4 mice per group). For abbreviations, see fig. S4. Mann-Whitney test, $*p < 0.05$. (**I**) Protocol for identification of stress-driven inputs to VTA^{Vgat} neurons that output to LH. (**J**) Immunostaining shows presynaptic inputs to LH-projecting VTA^{Vgat} neurons from LPO, and cFOS-positive cells activated by stress. Scale bar, 200 μm and 50 μm (inset). (**K**) Summary statistics of activated fractions to LH-projecting VTA^{Vgat} neurons (n = 4 mice per group). Mann-Whitney test, $*p < 0.05$. (**L**) Schematic diagram summarizing the stress-driven input-output relations.

cells (fig. S15A): A large proportion (42%) expressed *vgat*/somatostatin (*sst*), and others were characterized by *vgat*/parvalbumin (*pv*) (10%) or *vgat expression* alone (32%), the remaining cells being 2% *vgat/vglut2*, 2% *vgat/vglut2/sst*, and 2% *vgat/vip* (fig. S15B). We further characterized the SDS-activated cells using reporter mice (Fig. 4, A B and fig. S15C, D). Nearly 40% of the VTA^{Sst} neurons expressed cFOS after SDS (Fig. 4A and fig. S15, E and F), whereas there was no induction of cFOS following SDS in VTA^{Pv} cells (Fig. 4B). We next determined the activity of individual subtypes responding to stress using fiber photometry (Fig. 4, C and D). Both the VTA^{Sst} and VTA^{Pv} populations responded transiently to SDS, but the collective calcium signal for VTA^{Sst} cells was larger (Fig. 4, C and D), and only VTA^{Sst} neurons had persistent activation after SDS, with enhanced activity for a few hours (Fig. 4E). By contrast, the transient activity of VTA^{Pv} neurons after SDS was not sustained (Fig. 4F).

We next used activity tagging with hM3Dq to capture SDS-tagged VTA^{Sst} neurons (fig. S16, A and B). Because VTA^{Sst} cells are heterogeneous (*28*), we examined the molecular identities of SDS-tagged cells (fig. S16C). These tagged cells predominantly expressed *vgat/gad1* (90%) (fig. S16D).

We tested whether VTA^{Sst} cells could respond to two types of insomnia-inducing stress, restraint and cage change (*7*, *29*). However, these procedures did not affect the acute or long-term calcium activity in VTA^{Sst} neurons (fig. S17, A to D). In addition, we did not observe any VTA^{Sst} neurons becoming tagged by restraint stress or cage-change stress (fig. S17E).

Next, we measured spontaneous activities of VTA^{Sst} neurons across brain states. From calcium photometry, VTA^{Sst} neurons were primarily active during spontaneous NREM and REM sleep (Fig. 4, G and H), whereas VTA^{Pv} neurons were wake-active (Fig. 4, I and J). Chemogenetic stimulation of VTA^{Sst} neurons directly increased sleep (fig. S18). We further defined how VTA^{Sst} neurons link stress and sleep. We recorded the spontaneous activity of stress-tagged VTA^{Sst} populations across brain states (fig. S19A). The tagged cells were also primarily active during NREM and REM sleep (fig. S19B). Chemogenetic reactivation of SDS-tagged VTA^{Sst} cells was sufficient to promote NREM and REM sleep (fig. S19, C to E).

To explore whether the VTA^{Sst}→LH pathway links stress and sleep, we conducted fiber photometry to measure terminal activity in the LH that responded to stress by expressing GCaMP6 in VTA^{Sst} neurons (fig. S20A). The VTA^{Sst}→LH projection responded to SDS (fig. S20B). Following intersectional activity tagging during SDS (fig. S20C), chemogenetic reactivation of the VTA^{Sst}→LH pathway promoted sleep (fig. S20, D and E).

Finally, we examined directly whether VTA^{Sst} neurons are necessary for SDS-induced sleep. Genetic ablation specifically depleted VTA^{Sst} neurons (fig. S21). Lesioning of VTA^{Sst} neurons decreased baseline sleep (Fig. 5, A to C). When VTA^{Sst}-caspase mice were challenged with SDS, SDS-induced sleep was abolished (Fig. 5, B and C). This was also confirmed by chemogenetic manipulation; inhibition of VTA^{Sst} neurons also decreased SDS-induced sleep (fig. S22). By contrast, ablation of VTA^{Pv} neurons decreased baseline NREM sleep, but SDS-induced sleep could still be elicited (Fig. 5, D to F).

SDS-induced sleep by VTA^{Sst} neurons reduces stress-induced anxiety

Given the proposed restorative function of sleep after SDS (Fig. 1, D to G), we explored whether this function was linked to VTA$^{Vgat\text{-}Sst}$ neurons. Ablation of VTA^{Sst} neurons or chemogenetic inhibition of SDS-tagged VTA^{Vgat} neurons had no effect on baseline anxiety-like behaviors (fig. S23). However, after SDS, mice lacking SDS-induced sleep (because of selective lesioning or inhibition of VTA^{Sst} neurons or inhibition of SDS-tagged VTA^{Vgat} neurons (Figs. 2, F to H, and 5, A to C, and fig. S22) remained in an anxious state (Fig. 6, A to C and fig. S24), similar to the effects of sleep deprivation after SDS (Fig. 1, E and F, and fig. S25). When VTA$^{Vgat\text{-}Sst}$ neurons were unimpeded, while the mice had sufficient SDS-induced sleep, the SDS-induced anxiety-like behaviors were reduced to baseline (Fig. 6, B and C, and fig. S24). We found that sleep deprivation after SDS suppressed activity in VTA^{Sst} neurons induced by SDS (fig. S26). However, during the sleep deprivation procedure after SDS while mice were awake, VTA^{Sst} cell stimulation did not reduce anxiety even if VTA^{Sst} neurons were activated (fig. S27), suggesting that the anxiolytic effects require SDS-induced sleep.

Activation of VTA^{Sst} neurons suppresses corticotrophin-releasing factor levels induced by SDS

How do VTA^{Sst} neurons regulate CORT production? VTA^{Sst} neurons expressing hM3Dq-mCherry sent numerous mCherry-positive axons into the paraventricular nucleus (PVN) area (fig. S28, A and B), a major site of corticotrophin-releasing factor (CRF) production. After SDS, cells in the PVH are excited, as inferred from their strong expression of cFOS; but stimulation of VTA^{Sst} neurons inhibited SDS-activated cells in the PVN (fig. S28C).

We next used a genetically encoded CRF sensor (*AAV-hSyn-GRAB$_{CRF1.0}$*) (*30*) to determine the dynamics of CRF release around the PVN (fig. S28D and Fig. 6D). CRF sensor signals were indistinguishable between control and chemogenetic activation of VTA^{Sst} neurons (fig. S28E), consistent with CORT concentrations also not changing with VTA^{Sst} stimulation (fig. S28F). But after SDS, there were large increases in CRF (fig. S28G). However, chemogenetic activation of VTA^{Sst} neurons prevented this increase (Fig. 6E), consistent with correspondingly decreased CORT concentrations (Fig. 6F). By contrast, chemogenetic inhibition of VTA^{Sst} neurons further increased SDS-induced CRF concentrations (Fig. 6G), thereby increasing CORT levels after SDS (Fig. 6H).

SDS-induced sleep by VTA^{Sst} neurons reduces CORT concentrations

For mice unable to have SDS-induced sleep, either because their VTA^{Sst} neurons had been ablated or were inhibited (Fig. 6I and fig. S29A), CORT concentrations remained higher during

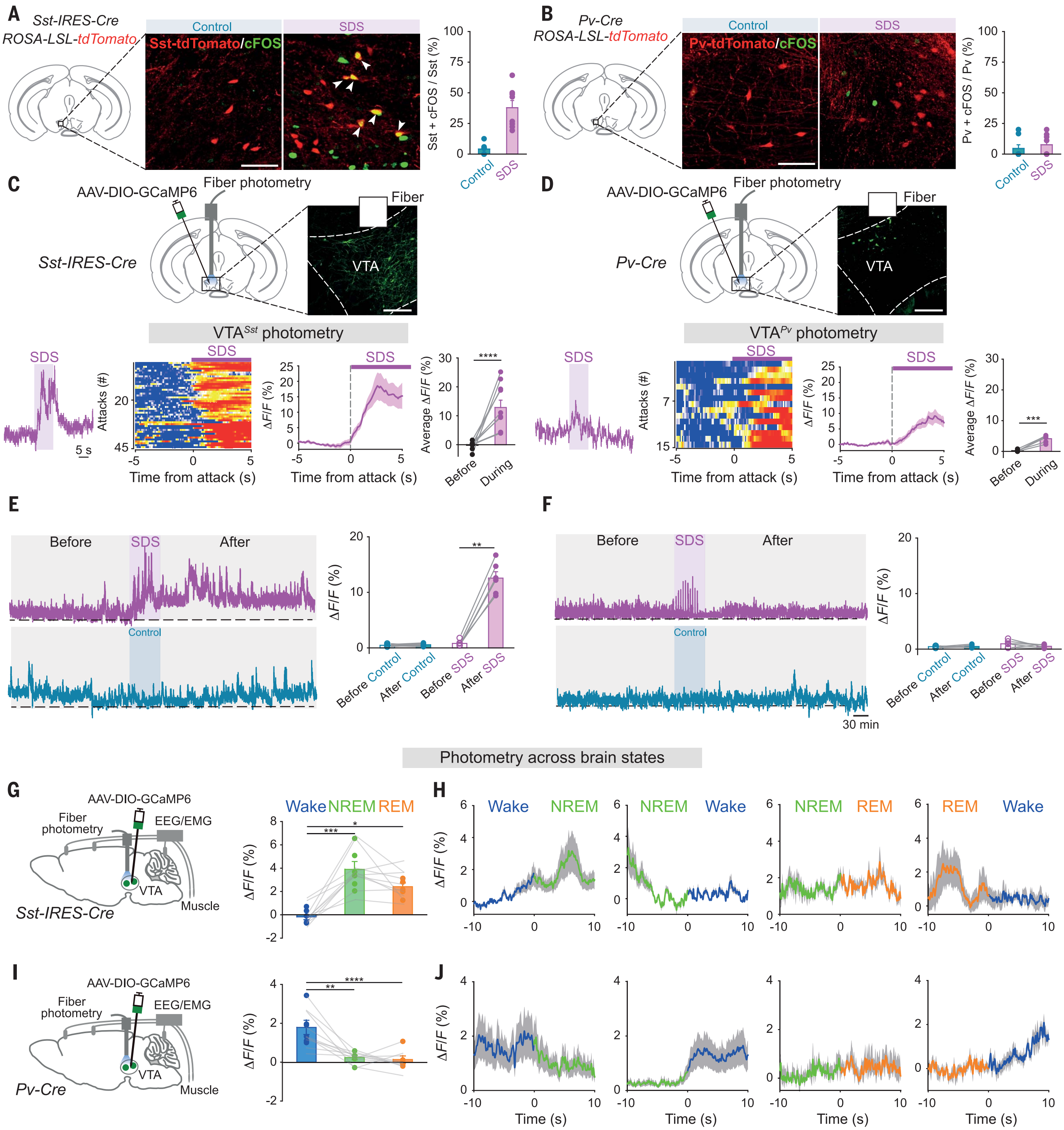


Fig. 4. Activity of VTA$^{Sst or Pv}$ neurons responding to stress and across brain states. (**A** and **B**) cFOS expression and quantification in genetically labeled VTA^{Sst} (A) or VTA^{Pv} (B) neurons after control experience or SDS (n = 8 mice per group). Scale bar, 100 μm. (**C** and **D**) Fiber photometry measuring calcium signals in VTA^{Sst} (n = 10 mice per group, 45 trials) (C) or VTA^{Pv} (n = 6 mice per group, 15 trials) (D) neurons responding to SDS. Raw calcium signal traces, color matrix of signals for all trials, $\Delta F/F$ ratios across the experimental period and average $\Delta F/F$ ratios before and during SDS. Paired t test, ***$p < 0.001$, ****$p < 0.0001$. Scale bar, 200 μm. (**E** and **F**) Fiber photometry measuring long-term calcium signals in VTA^{Sst} (E) or VTA^{Pv} (F) neurons. Raw traces and average $\Delta F/F$ ratios before and after the procedures (n = 6 mice per group). Paired t test, **$p < 0.01$. (**G** to **J**) Fiber photometry with electroencephalography and electromyography measuring spontaneous activity across brain states. $\Delta F/F$ ratios in VTA^{Sst} (G) or VTA^{Pv} (I) neurons during wakefulness, NREM and REM sleep, and at transitions of vigilance states (n = 6 mice per group) (H and J). One-way repeated ANOVA, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

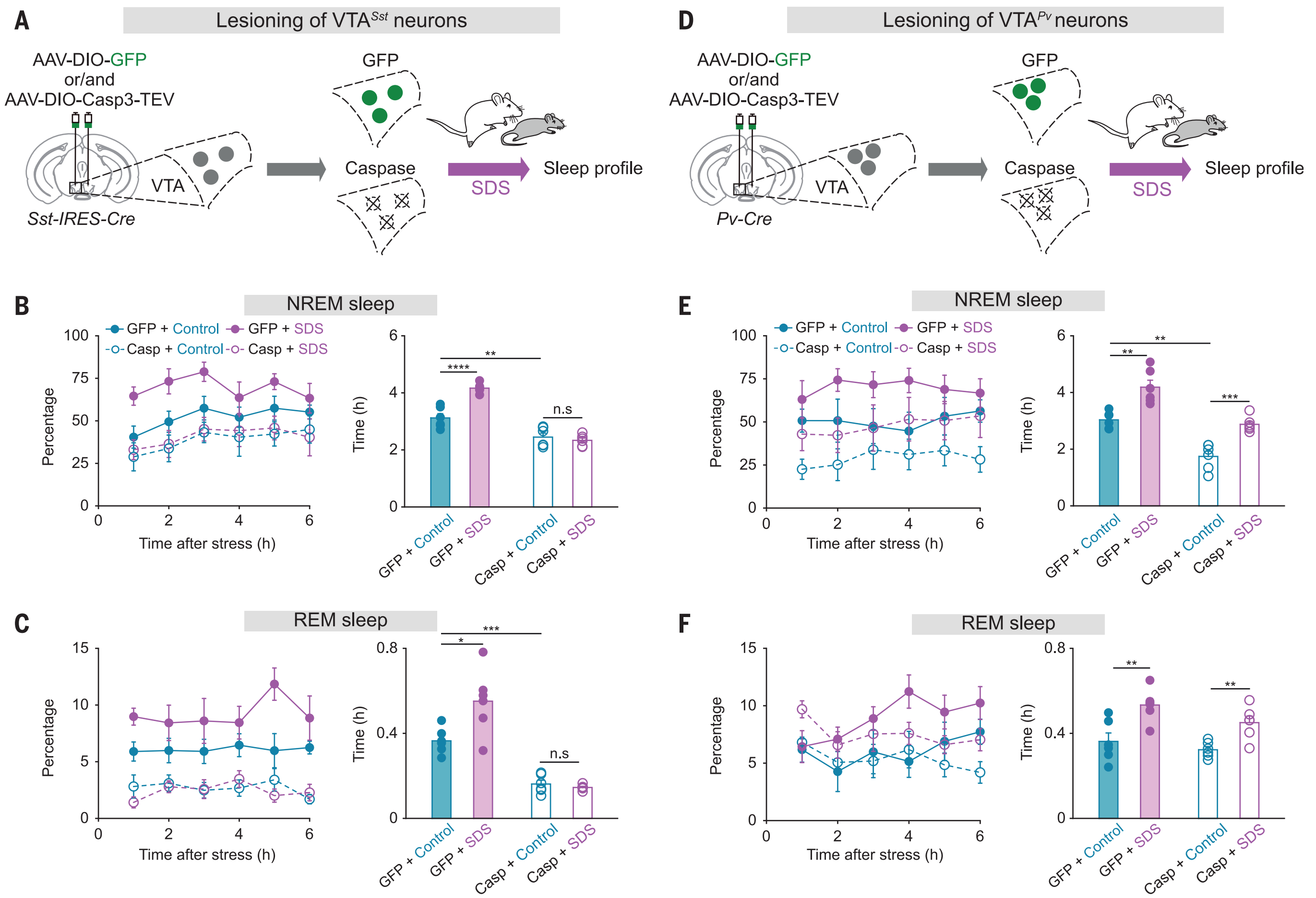


Fig. 5. VTA^{Sst} neurons are necessary for SDS-induced sleep. (**A** and **D**) Genetic ablation of VTA^{Sst} (A) or VTA^{Pv} (D) neurons. (**B**, **C**, **E**, **F**) Percentage and time of NREM or REM sleep in VTA^{Sst} (B and C) or VTA^{Pv} (E and F) ablated mice or control mice given control or SDS. Two-way ANOVA with bonferroni post hoc test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, **** $p < 0.0001$; n.s., not significant.

their home cage sleep after SDS (Fig. 6J and fig. S29B), similar to the effects of sleep deprivation after SDS (Fig. 1G). However, when VTA$^{Vgat\text{-}Sst}$ neurons were unimpeded, the SDS-induced sleep correlated with CORT concentrations returning to baseline (Fig. 6J and fig. S29B). In addition, activation of VTA^{Sst} cells during sleep deprivation after SDS, (i.e., activation of these cells while mice were awake) partially reduced CORT concentrations (fig. S29, C and D), but the overall CORT levels still remained elevated (fig. S29D), suggesting that sleep after SDS is also needed to reduce CORT concentrations.

Discussion

Our proposed circuit model for how SDS translates to sleep and reduction of anxiety, with VTA$^{Vgat\text{-}Sst}$ cells playing a central role, is shown in Fig. 6K. Once activated by SDS, VTA$^{Vgat\text{-}Sst}$ cells drive sleep through the LH, a brain region containing a diverse population of cells implicated in regulating stress, anxiety, and sleep-wake behaviors (*31*, *32*). VTA$^{Vgat\text{-}Sst}$ cell activity is maintained for some hours beyond the stress episode, suggesting a form of plasticity that enables them to keep promoting NREM and REM sleep episodes for a sustained period. In parallel to their sleep-inducing and anxiety-reducing effects, VTA$^{Vgat\text{-}Sst}$ cells inhibit CRF-producing neurons in the PVN hypothalamus, thereby reducing CORT concentrations after SDS. We found that SDS-induced anxiety persisted even in the presence of CORT inhibitors. These results suggest that physiological activation of VTA^{Sst} neurons during and after SDS represses CRF and therefore CORT production, guarding against overproduction of CORT. Persistently increased CORT concentrations have deleterious effects on body organs (*1*). We propose that the reduced anxiety comes from the sleep component. After SDS, the restorative sleep by VTA$^{Vgat\text{-}Sst}$ cells also aids CORT concentrations returning to baseline, so there seem to be parallel routes to reducing CORT levels, but with VTA^{Sst} cells coordinating both mechanisms.

The output pathways regulated by VTA$^{Vgat\text{-}Sst}$ cells in the LH to induce sleep and reduce anxiety are unclear. VTA^{Gad67} neurons inhibit orexin/Hcrt neurons in the LH (*20*). However, chemogenetic inhibition of LH^{Hcrt} cells did not reduce anxiety after SDS (fig. S30), and orexin receptor antagonists did not restore the anxiolytic effects that were missing in VTA^{Sst}-lesioned mice that had undergone SDS (fig. S31), suggesting that orexin/Hcrt cell inhibition is not required for the anxiolytic actions of VTA$^{Vgat\text{-}Sst}$ cells. Thus, identifying the targets of VTA$^{Vgat\text{-}Sst}$ cells requires further study. Local action within the VTA of the VTA^{Sst} neurons is also possible.

We have shown here that GABA^{Sst} neurons in the VTA respond to SDS, an ethological model for psychosocial stress, by inducing restorative sleep and decreasing CRF production. Targeting these neurons could potentially provide a new route for treating anxiety disorders.

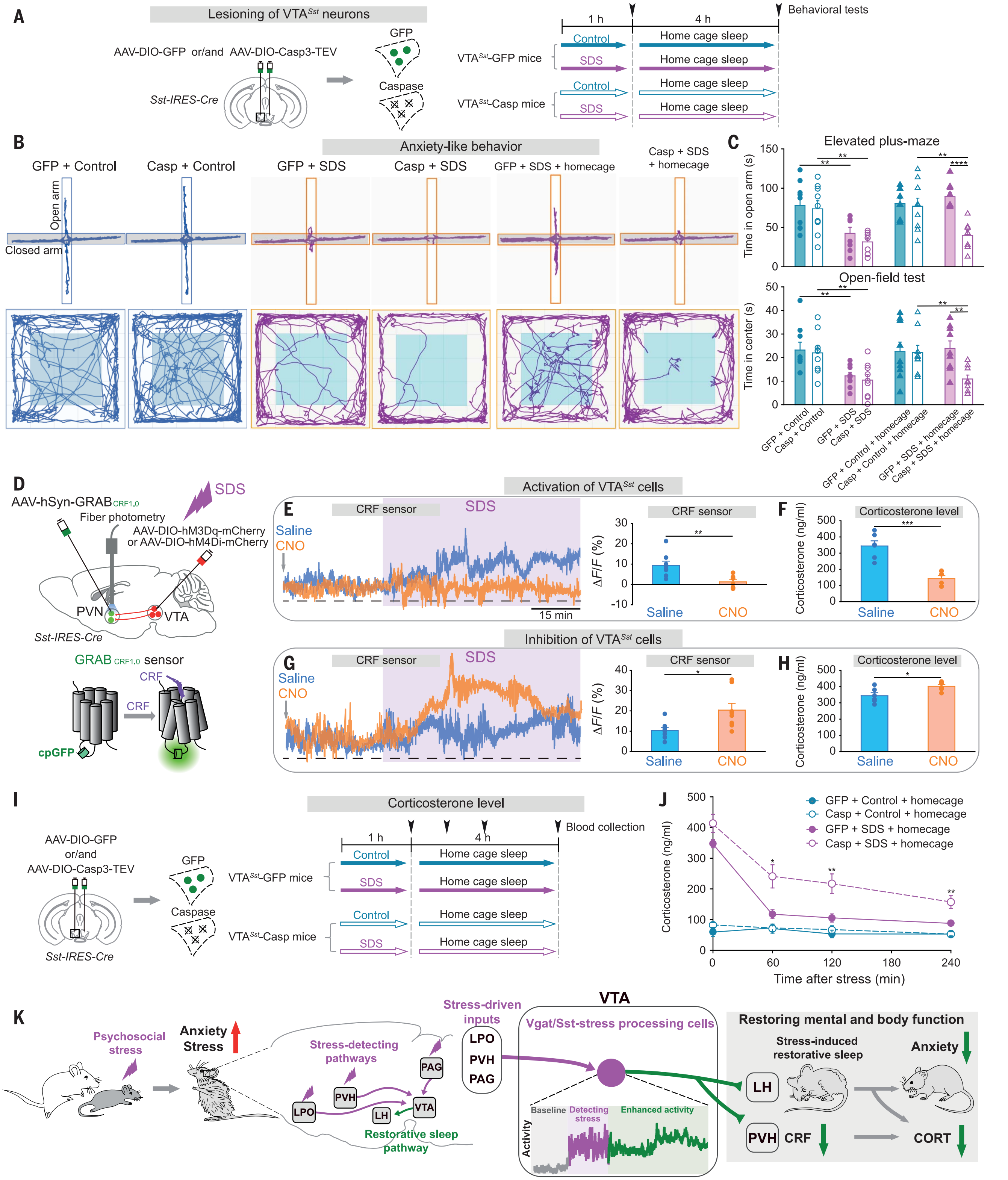


Fig. 6. Activity and SDS-induced sleep by VTA^{Sst} neurons reduces anxiety and corticosterone concentrations. (**A** to **C**) Plan of the experimental procedure (A), tracing of locomotion for representative animals (B), time spent in the open arms of the elevated plus maze and in the center zone during the open-field test (n = 9 mice per group) (C). (**D**) Delivering genetically encoded CRF sensor in the PVN hypothalamus alongside chemogenetic manipulation of VTA^{Sst} neurons. (**E** and **G**) Raw PVN CRF sensor traces and $\Delta F/F$ ratios during SDS after chemogenetic activation (E) or inhibition (G) of VTA^{Sst} neurons (n = 8 mice per group). (**F** and **H**) Corticosterone concentrations following SDS after chemogenetic activation (F) or inhibition (H) of VTA^{Sst} neurons (n = 6 mice per group). (**I** and **J**) Plan of the procedure (I) and corticosterone concentrations (n = 4 to 9 mice per group) (J). (**K**) Conceptual summary diagram. (C and J) Two-way ANOVA with bonferroni post hoc test, $*p < 0.05$, $**p < 0.01$, $****p < 0.0001$. (E to H) Unpaired t test, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

REFERENCES AND NOTES

1. G. Russell, S. Lightman, *Nat. Rev. Endocrinol.* **15**, 525–534 (2019).
2. B. S. McEwen, H. Akil, *J. Neurosci.* **40**, 12–21 (2020).
3. M. Nollet, W. Wisden, N. P. Franks, *Interface Focus* **10**, 20190092 (2020).
4. A. N. Goldstein, M. P. Walker, *Annu. Rev. Clin. Psychol.* **10**, 679–708 (2014).
5. D. A. Kalmbach, J. R. Anderson, C. L. Drake, *J. Sleep Res.* **27**, e12710 (2018).
6. E. J. W. Van Someren, *Physiol. Rev.* **101**, 995–1046 (2021).
7. S. B. Li *et al.*, *Sci. Adv.* **6**, eabc2590 (2020).
8. R. G. Foster, *Interface Focus* **10**, 20190098 (2020).
9. M. Nollet *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2733–2742 (2019).
10. N. P. Franks, W. Wisden, *Science* **374**, 556–559 (2021).
11. X. Feng *et al.*, *Neurosci. Bull.* **36**, 1137–1146 (2020).
12. N. Gujar, S. A. McDonald, M. Nishida, M. P. Walker, *Cereb. Cortex* **21**, 115–123 (2011).
13. M. Morales, E. B. Margolis, *Nat. Rev. Neurosci.* **18**, 73–85 (2017).
14. V. Krishnan *et al.*, *Nature* **543**, 507–512 (2017).
15. L. A. Gunaydin *et al.*, *Cell* **157**, 1535–1551 (2014).
16. C. Bouarab, B. Thompson, A. M. Polter, *Front. Neural Circuits* **13**, 78 (2019).
17. J. H. Jennings *et al.*, *Nature* **496**, 224–228 (2013).
18. X. Yu *et al.*, *Nat. Neurosci.* **22**, 106–119 (2019).
19. A. Eban-Rothschild, G. Rothschild, W. J. Giardino, J. R. Jones, L. de Lecea, *Nat. Neurosci.* **19**, 1356–1366 (2016).
20. S. Chowdhury *et al.*, *eLife* **8**, e44928 (2019).
21. Y. Takata *et al.*, *J. Neurosci.* **38**, 10080–10092 (2018).
22. Z. Zhou *et al.*, *Neuron* **103**, 473–488.e6 (2019).
23. A. Eban-Rothschild *et al.*, *eNeuro* **7**, ENEURO.0356-19.2020 (2020).
24. K. R. Tan *et al.*, *Neuron* **73**, 1173–1183 (2012).
25. S. Fujii, M. K. Kaushik, X. Zhou, M. Korkutata, M. Lazarus, *Front. Neurosci.* **13**, 322 (2019).
26. L. Faget *et al.*, *Cell Rep.* **15**, 2796–2808 (2016).
27. E. J. Paul, K. Tossell, M. A. Ungless, *Eur. J. Neurosci.* **50**, 3732–3749 (2019).
28. E. Nagaeva *et al.*, *eLife* **9**, e59328 (2020).
29. M. K. Kaushik, K. Aritake, A. Takeuchi, M. Yanagisawa, Y. Urade, *Sci. Rep.* **7**, 8892 (2017).
30. H. Wang *et al.*, A toolkit of highly selective and sensitive genetically encoded neuropeptide sensors. bioRxiv 10.1101/2022.03.26.485911 (2022).
31. E. Arrigoni, M. J. S. Chee, P. M. Fuller, *Neuropharmacology* **154**, 34–49 (2019).
32. S. B. Li *et al.*, *Science* **375**, eabh3021 (2022).

ACKNOWLEDGMENTS

We thank K. Beier, L. Q. Luo, and R. C. Malenka for sharing plasmids; A. Murray for providing rabies virus; Z. A. Hu for providing *Hcrt-IRES-Cre* mice; and J. Hu for providing *Sst-IRES-Cre* mice. **Funding:** Our work was supported by the Wellcome Trust (107839/Z/15/Z, 107841/Z/15/Z, and 220759/Z/20/Z, to N.P.F. and W.W.); the UK Dementia Research Institute (UK DRI-5004, to W.W. and N.P.F.); and the National Natural Science Foundation of China (grant nos. 82030038, 81620108012, and 81901080, to H.D. and N.P.F.). **Author contributions:** X.Y., N.P.F., and W.W. conceived—and with G.Z. and H.D.—designed the experiments; X.Y., G.Z., D.W., S.W., A.L., R.L., H.L., J.Z., J.L., M.N., Y.C., T.Z., M.C., and R.Y. performed the experiments and/or data analysis; A.L.V. provided the Neurologgers; H.W. and Y.L. engineered the CRF sensor; N.P.F. and W.W. contributed to the data analysis and with H.D. supervised the project; X.Y., N.P.F., and W.W. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data necessary to understand and assess the conclusions of this study are available in the manuscript or the supplementary materials. Constructs generated in this study have been deposited at Addgene with accession numbers provided in the supplementary materials.
License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn0853
Materials and Methods
Figs. S1 to S33
References (*33*–*59*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 2 November 2021; accepted 6 May 2022
10.1126/science.abn0853

HUMAN GENOMICS

Ancient DNA reveals five streams of migration into Micronesia and matrilocality in early Pacific seafarers

Yue-Chen Liu[1,2]*, Rosalind Hunter-Anderson[3]*, Olivia Cheronet[4], Joanne Eakin[5], Frank Camacho[6], Michael Pietrusewsky[7], Nadin Rohland[1,8], Alexander Ioannidis[9,10], J. Stephen Athens[11], Michele Toomay Douglas[11], Rona Michi Ikehara-Quebral[11], Rebecca Bernardos[1], Brendan J. Culleton[12], Matthew Mah[1,8,13], Nicole Adamski[1,13], Nasreen Broomandkhoshbacht[1,13], Kimberly Callan[1,13], Ann Marie Lawson[1,13], Kirsten Mandl[4], Megan Michel[1,13], Jonas Oppenheimer[1,13], Kristin Stewardson[1,13], Fatma Zalzala[1,13], Kenneth Kidd[14], Judith Kidd[14], Theodore G. Schurr[15], Kathryn Auckland[16], Adrian V. S. Hill[16,17], Alexander J. Mentzer[16,18], Consuelo D. Quinto-Cortés[19], Kathryn Robson[20], Douglas J. Kennett[21], Nick Patterson[2,8], Carlos D. Bustamante[10,22]†, Andrés Moreno-Estrada[19], Matthew Spriggs[23,24], Miguel Vilar[25], Mark Lipson[1,2], Ron Pinhasi[4,26]*, David Reich[1,2,8,13]*

Micronesia began to be peopled earlier than other parts of Remote Oceania, but the origins of its inhabitants remain unclear. We generated genome-wide data from 164 ancient and 112 modern individuals. Analysis reveals five migratory streams into Micronesia. Three are East Asian related, one is Polynesian, and a fifth is a Papuan source related to mainland New Guineans that is different from the New Britain–related Papuan source for southwest Pacific populations but is similarly derived from male migrants ~2500 to 2000 years ago. People of the Mariana Archipelago may derive all of their precolonial ancestry from East Asian sources, making them the only Remote Oceanians without Papuan ancestry. Female-inherited mitochondrial DNA was highly differentiated across early Remote Oceanian communities but homogeneous within, implying matrilocal practices whereby women almost never raised their children in communities different from the ones in which they grew up.

Modern humans arrived in Near Oceania at least 47,000 years before present (BP) and spread through Australia, New Guinea, the Bismarck Archipelago, and the Solomon Islands (*1*, *2*). After 3500 to 3300 BP, humans expanded into previously unoccupied Remote Oceania (Fig. 1A).

In the southwest Pacific, the earliest archaeological sites are associated with artifacts of the Lapita complex, appearing in the Bismarck Archipelago as early as ~3350 BP and reaching the unoccupied islands of Remote Oceania by 3000 to 2850 BP (*3*, *4*). Ancient DNA from 11 individuals from Vanuatu and Tonga 3000 to 2500 BP indicates that these pioneers were related distantly to Neolithic southeastern Chinese (*5*), more closely related to Neolithic and Iron Age people of Taiwan (*6*), and most closely related to the ancestors of present-day north-central Philippine groups such as Kankanaey Igorot (*7*–*10*). However, the primary ancestry of many southwest Pacific Islanders today is "Papuan" (our term to describe the primary ancestry of peoples of New Guinea, the Bismarck Archipelago, and the Solomon Islands), which genetic data has shown is due to a secondary expansion that began ~2500 BP (*7*–*10*).

The first people to reach the Mariana Archipelago arrived around 3500 to 3200 BP (*11*–*14*). Their material culture (*15*) differed markedly from the Lapita assemblages in the southwest Pacific, with Marianas Redware ceramics being more similar to those found at sites in the Philippines and at the northern tip of Sulawesi (*16*). This study uses a revised chronology for the archaeology of the Mariana Islands that terms the earliest three periods of occupation from 3500 to 1600 BP "Unai" (table S1). The burials that we analyze date to 2800 to 2200 BP (Middle to Late Unai) and thus may not reflect the ancestry profile of Early Unai inhabitants. After 1100 BP, distinctive megaliths (latte) began to appear in the Mariana Islands, along with other material cultural changes marking the "Latte" period. The oldest evidence of human occupation in Palau in Western Micronesia dates to ~3000 BP (*17*). The oldest evidence in Central Micronesia is ~2000 BP; ceramics at these sites are similar to late Lapita pottery and shell artifacts and thus could reflect roots in earlier Lapita cultures in either northern New Guinea or in the southwest Pacific (*18*, *19*).

Linguistic relationships among Malayo-Polynesian (MP) languages that comprise all Austronesian languages outside of Taiwan provide an independent source of information about the cultural and geographic origins of Micronesian peoples (fig. S1). The CHamoru (*20*) language spoken by the indigenous people of the Mariana Islands is a first-order branch within MP; Palauan is another. All other Micronesian languages and languages of the southwest Pacific and Polynesia comprise a third major branch, Central-Eastern Malayo-Polynesian (CEMP) (*21*–*23*). Most Micronesian CEMP languages form a Nuclear Micronesian subgroup, which has been hypothesized to have developed somewhere between the Admiralty Islands and Vanuatu and to have spread near the end of the Lapita period ~2500 BP (*24*). By contrast, Yap's language is believed to be an early offshoot of Proto-Oceanic derived directly from protolanguages that branched during the Lapita expansion, although Yapese was also subsequently affected by borrowings from other languages (*25*). The people of Kapingamarangi and Nukuoro atolls in the Caroline Islands speak Polynesian languages, suggesting replacement of the original languages by Polynesian immigration (*26*, *27*).

To test alternative models of population history, we generated genome-wide ancient DNA data for 164 individuals from five archaeological sites and coanalyzed them with published data from two ~2200 BP individuals from

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [3]Independent Researcher, Albuquerque, NM 87106, USA. [4]Department of Evolutionary Anthropology, University of Vienna, Vienna 1030, Austria. [5]Independent Researcher, Albuquerque, NM 87107, USA. [6]Department of Biology, University of Guam, Mangilao 96923, Guam. [7]Department of Anthropology, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA. [8]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [9]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA. [10]Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA. [11]International Archaeological Research Institute, Inc., Honolulu, HI 96826, USA. [12]Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA 16802, USA. [13]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. [14]Department of Genetics, Yale Medical School, New Haven, CT 06520, USA. [15]Department of Anthropology, University of Pennsylvania, Philadelphia, PA 19104, USA. [16]Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [17]The Jenner Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7DQ, UK. [18]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, UK. [19]National Laboratory of Genomics for Biodiversity (LANGEBIO), Unit of Advanced Genomics, CINVESTAV, Irapuato 36821, Mexico. [20]MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. [21]Department of Anthropology, University of California, Santa Barbara, CA 93106, USA. [22]Center for Computational, Evolutionary and Human Genomics (CEHG), Stanford University, Stanford, CA 94305, USA. [23]School of Archaeology and Anthropology, The Australian National University, Canberra, ACT 2601, Australia. [24]Vanuatu National Museum, Vanuatu Culture Centre, PO Box 184, Port Vila, Vanuatu. [25]Department of Anthropology, University of Maryland, College Park, MD 20742, USA. [26]Human Evolution and Archaeological Sciences, University of Vienna, 1030 Vienna, Austria.
***Corresponding author. Email: yuechen_liu@hms.harvard.edu (Y.-C.L.); rozinabq@gmail.com (R.H.-A.); ron.pinhasi@univie.ac.at (R.P.); reich@genetics.med.harvard.edu (D.R.)**
†Present address: Galatea Bio, Inc., Hialeah, FL 33010, USA.

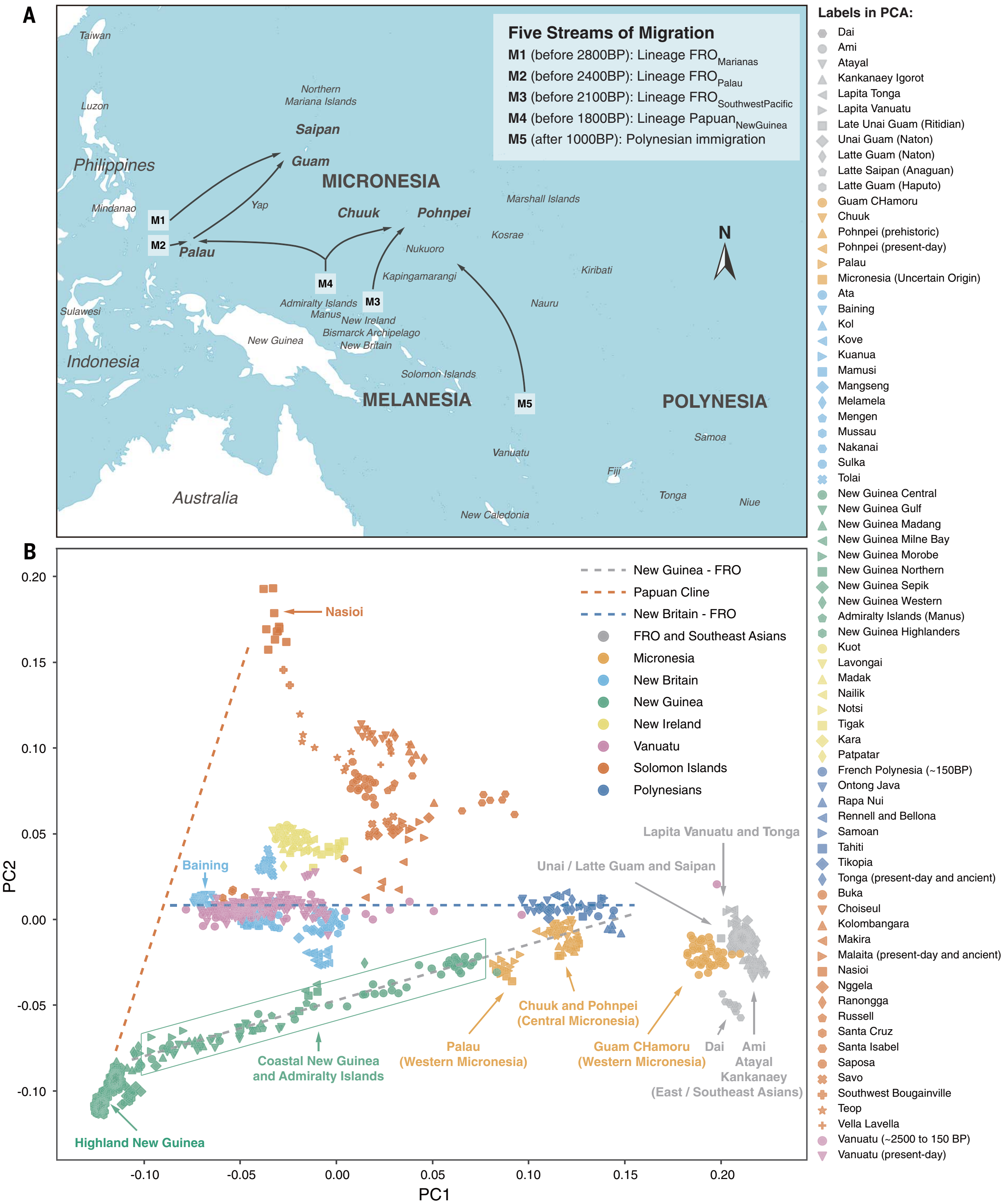


Fig. 1. Map and PCA. (**A**) Map showing five inferred streams of migration into Micronesia. (**B**) PCA results. Axes are computed with Dai, Nasioi, and Papuans; others are projected.

Guam (*28*). A total of 109 individuals (2800 to 300 BP) were from the Unai and Latte periods in Guam, 46 (600 to 200 BP) from the Latte period in Saipan, and 11 (500 to 300 BP) from Na Island and the nearby Nan Madol site in Pohnpei's protected lagoon in Central Micronesia (*20*).

We prepared samples in clean rooms, extracted DNA, built sequencing libraries, enriched for a common panel of ~1.2 million single-nucleotide polymorphisms (SNPs), and sequenced them (*20*). For individuals with evidence of high contamination, we restricted analysis to sequences with evidence of characteristic ancient DNA damage (*20*). The analyzed individuals had a median of 558,971 SNPs with data (table S2). We also genotyped 112 present-day Micronesians mainly from Guam, Palau, Chuuk, and Pohnpei (tables S3 and S4). We obtained 31 direct radiocarbon dates, 30 of which were on the same samples we analyzed

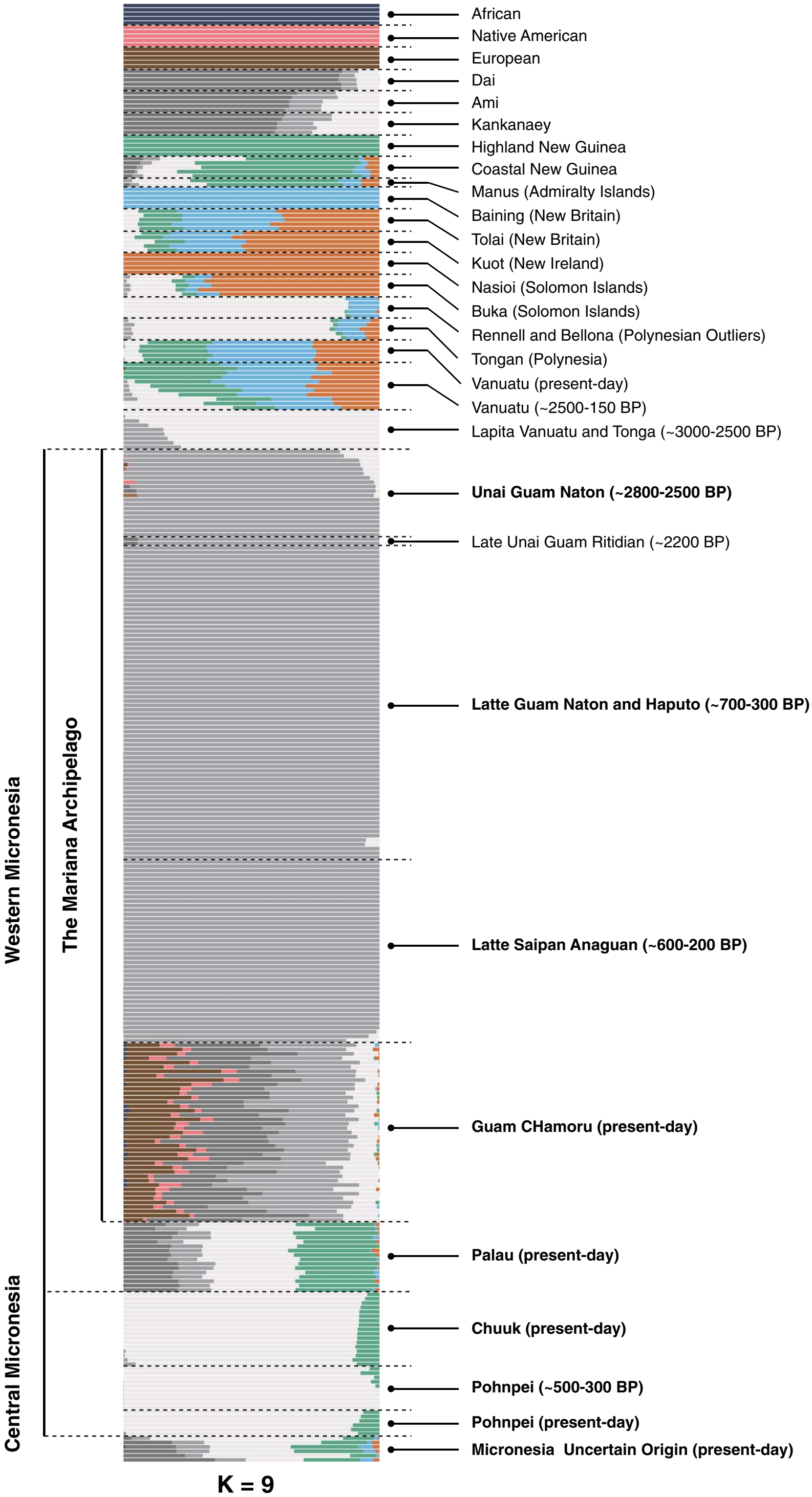


Fig. 2. Clustering analysis. Unsupervised ADMIXTURE (K = 9 clusters). New data are in boldface.

for DNA (tables S5 and S6). We coanalyzed our newly produced data with published data from 95 prehistoric individuals and 1642 present-day individuals from globally diverse populations (table S7).

Overview of population structure

We carried out principal components analysis (PCA) (Fig. 1B and figs. S2 and S3) by computing axes using shotgun data of present-day Dai (southern China), Nasioi (Solomon Islands), and New Guineans (from the Eastern Highlands and Middle Sepik areas) and then projecting other individuals. The first principal component (PC) corresponds to the proportion of East Asian–associated ancestry, henceforth "First Remote Oceanian (FRO)" (PC1; lower on left, higher on right); the second PC differentiates Papuan ancestry from the Solomon Islands to New Guinea (PC2; up to down). The Unai, Latte, and Lapita individuals cluster with present-day people from the Philippines (Kankanaey) and Taiwan (Ami and Atayal) on the right, corresponding to high East Asian–associated ancestry. Two clines are visible. The first (dashed blue) links groups with high proportions of FRO ancestry to New Britain, Vanuatu, and Polynesia; the second (dashed gray) links to groups from New Guinea, the Admiralty Islands, Palau, and a genetically homogeneous group of Central Micronesians (Chuuk, Pohnpei, and prehistoric Pohnpei). This suggests admixture in variable proportions between FRO and Papuan ancestry from at least two different sources—more related to New Britain in the first case and New Guinea in the second. f_3-statistics reveal patterns qualitatively similar to those shown in the PCA (fig. S4 and table S8).

We also computed the symmetry statistic f_4 (X, Kankanaey Igorot; New Guinea Highlanders, Dai) to test which individuals had significant Papuan admixture (using Kankanaey as a baseline with no evidence of Papuan ancestry) (table S9). Unai and Latte individuals had little or no Papuan ancestry; except for four Latte individuals, we observed non-significant Z-tests based on the normally distributed score being $|Z| < 3$ standard errors from zero. Lapita individuals from Vanuatu and Tonga had a small, but nonzero, proportion of Papuan ancestry (0.4 to 4.4% and 3.3 to 7.7%, respectively) (*7–10*). Papuan admixture was present in all prehistoric and present-day individuals from Pohnpei (~27%) and all present-day people from Chuuk (~27%) and Palau (~38%). In modern CHamoru, the inferred Papuan ancestry is consistent with zero, making CHamoru the only genetically analyzed indigenous Remote Oceanian group without evidence of such ancestry.

Unsupervised clustering using ADMIXTURE recapitulates the patterns in the PCA and differentiates the FRO components of First Remote Oceanians (we show K = 9 clusters in Fig. 2; see also figs. S5 to S8). Two clusters correspond to East Asian–associated ancestry, with a light gray component maximized in Lapita individuals and a dark gray component maximized in Mariana individuals. Pohnpei and Chuuk in Central Micronesia primarily have a light gray Lapita-associated component. Modern CHamoru of Guam is the population with the highest proportion of dark gray, suggesting local continuity. Palau and Central Micronesia only have the green Papuan-associated

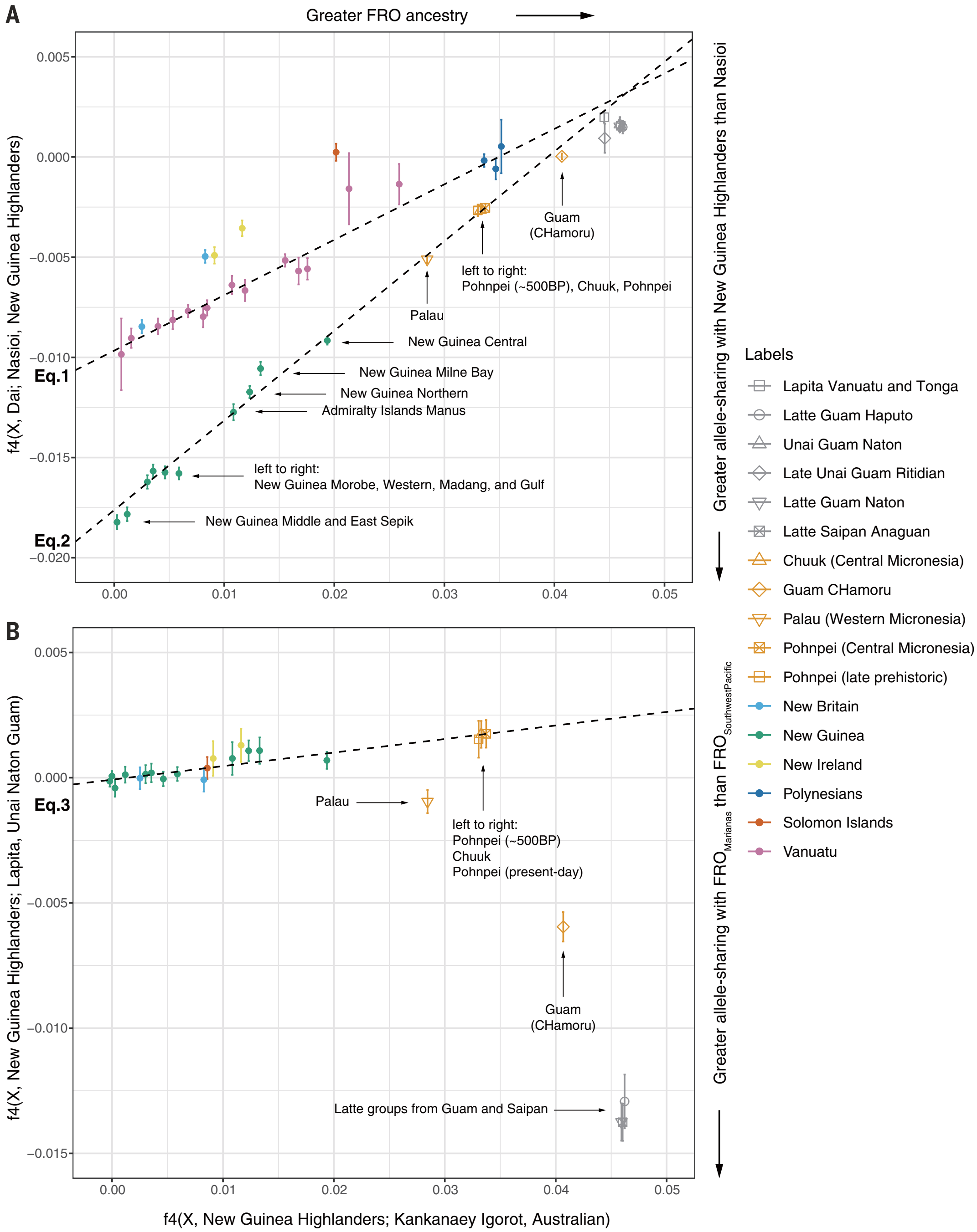


Fig. 3. Different Papuan and East Asian affinities. (**A** and **B**) Test for differential (A) Papuan and (B) FRO affinities using a merge of the 1240K and MEGA data (~169,000 SNPs). Equation 1 (Eq. 1) is computed with all groups from Vanuatu and Polynesians, Eq. 2 with all Micronesian and New Guinea–related groups except those from Guam and Saipan, and Eq. 3 with all present-day groups except Micronesians. We show one standard error in each direction on the *y* axis. We merged Lapita individuals from Vanuatu and Tonga. See fig. S9 for the same analysis performed on individuals for whom we have ~397,000 SNPs genotyped on a merge of 1240K and Human Origins data.

component maximized in New Guinea, without the orange-blue-green mixture characteristic of New Britain, the southwest Pacific, and Polynesia, suggesting previously undocumented Papuan spreads into Micronesia.

Evidence for at least five streams of migration into Micronesia

To determine the minimum number of migration streams into Micronesia needed to explain the data, we computed a statistic f_4(X, New Guinea Highlanders; Kankanaey Igorot, Australian) proportional to FRO ancestry and correlated it to statistics sensitive to different types of East Asian and Papuan-associated ancestry (*9*). We identified at least five distinct migratory streams, as follows.

(M1 to M3) Three streams of FRO migration into Micronesia including a previously unknown lineage. We plotted a statistic measuring affinity to the two previously identified (*7*, *28*) lineages $FRO_{SouthwestPacific}$ and $FRO_{Marianas}$, specifically, f_4(X, New Guinea Highlanders; Lapita, Unai) against our statistic measuring overall FRO ancestry proportion. All populations from the southwest Pacific and Polynesia fall on a line with a positive slope, implying closer affinity to Lapita than to Unai consistent with the Lapita-associated lineage being the source of their East Asian–associated ancestry (all residuals $|Z| < 2$ after regression; Fig. 3B and fig. S9B). Individuals from Central Micronesia (Pohnpei and Chuuk, and some other present-day Micronesians) also closely track the line (all residuals $|Z| < 2$), suggesting FRO ancestry from the Lapita expansion. By contrast, present-day individuals from Palau and the Mariana Islands yield negative f_4-statistics (all residuals $|Z| > 4$), implying FRO

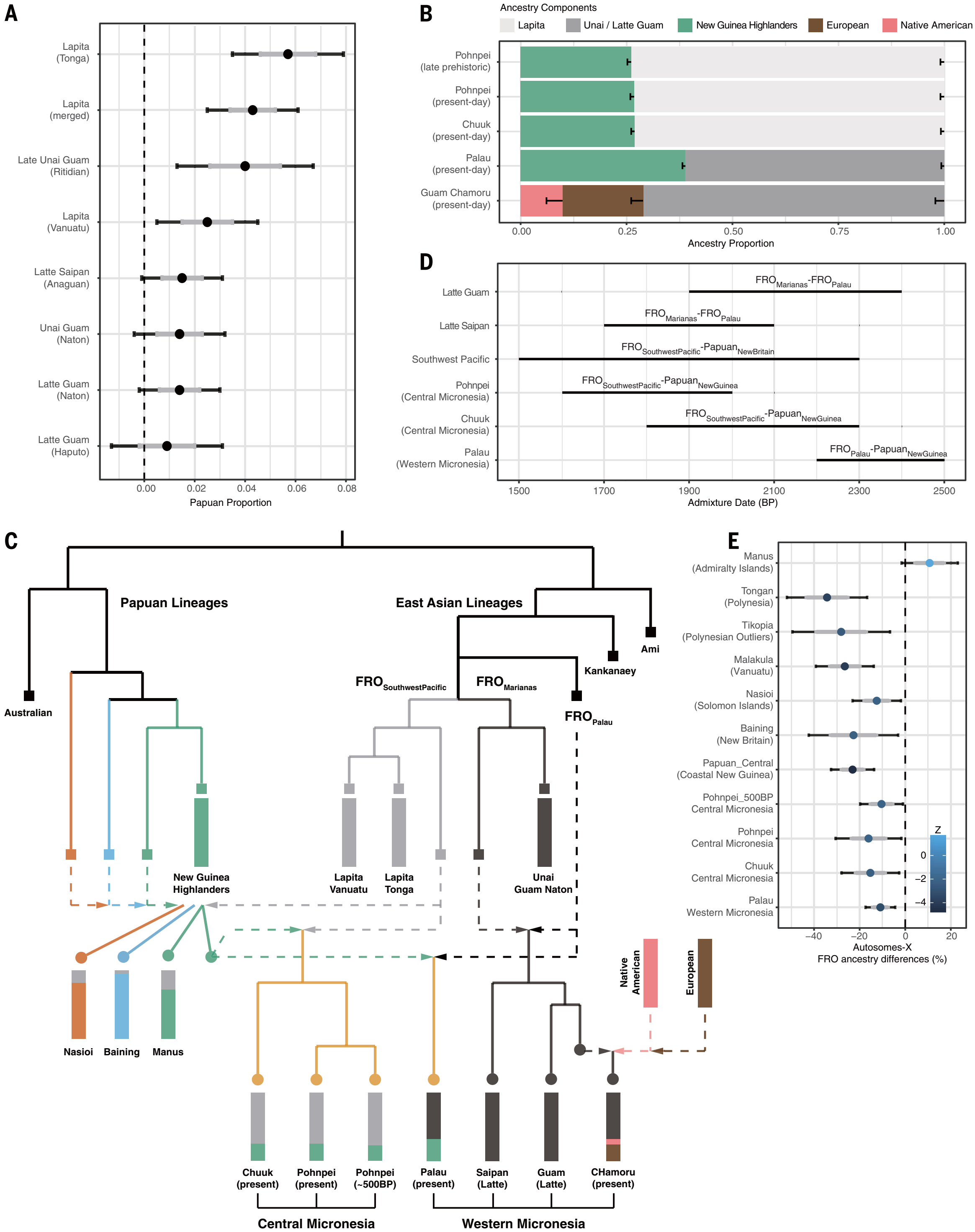


Fig. 4. Quantification of admixture events. (**A**) Proportions of Papuan ancestry in FRO and Latte groups. Thick and thin error bars show one standard error and 95% confidence interval, respectively. (**B**) Ancestry proportions from *qpAdm*. Each group is represented by a horizontal bar and partitioned into colored segments, representing different sources of their ancestry. Error bars show one standard error. (**C**) Admixture graphs. Arrow pairs (head to head) denote admixture events. The heights of the colored bars give mixture proportions. (**D**) Date of admixture. Ranges show one standard error in each direction. (**E**) Difference between FRO ancestry estimates on the autosomes and the X chromosome.

sources less closely related to the Lapita individuals (tables S12 and S13). We confirmed with f_4-symmetry statistics that all the prehistoric Remote Oceanian groups with nearly entirely East Asian–associated ancestry (Lapita, Unai, and Latte) descended from a common ancestral FRO population (table S22), which split earlier from the ancestors of indigenous and Iron Age Taiwanese and even earlier from those of Kankanaey Igorot. A surprise is that despite the fact that the Latte and Unai individuals share more alleles with each other than either group does with Lapita, there is not a simple tree relating these three groups, with the statistic f_4(Latte, Unai; Lapita, diverse East Asians) yielding many significant negative Z values (maximum $|Z|$ > 4; table S26). This suggests that the Latte individuals harbor admixture from a basal FRO lineage, which split from the lineages ancestral to Unai and Lapita before they separated from each other, a scenario that fits the data in explicit admixture graph modeling (Fig. 4C and figs. S12 to S15). We call this third lineage FRO_{Palau}

because the proportion of this lineage is maximized in modern Palauans (where we estimate that it contributes 62% ancestry versus 15% in Latte individuals) (fig. S13A).

(M4) A previously unknown stream of Papuan migration into Micronesia. We computed f_4(X, Dai; Nasioi, New Guinea Highlanders), where the latter two populations are differentiated Papuan groups, and plotted it against our statistic measuring FRO proportion. Modern and prehistoric groups from the southwest Pacific and Polynesia fall on a line that also includes New Britain (all residuals $|Z| < 2$; Fig. 3A and fig. S9A), consistent with ancestry from a New Britain–associated source we call $\text{Papuan}_{\text{NewBritain}}$ (*8*–*10*). By contrast, all prehistoric and present-day individuals from Micronesia with evidence of Papuan ancestry fall below the line (all residuals $|Z| > 4$), mirroring the two-cline pattern in the PCA (tables S10 and S11). When we fit a separate line for Micronesians, New Guinea, and the Admiralty Islands, we observe no outliers with $|Z| < 2$, consistent with a previously unknown spread of Papuan ancestry from a lineage $\text{Papuan}_{\text{NewGuinea}}$ more closely related to New Guinea and the Admiralty Islands on its northern fringe.

(M5) Polynesian gene flow into Micronesia. We computed f_4(X, Tolai; Kankanaey Igorot, diverse Polynesians) (tables S14 to S20), and plotted it against our f_4-statistic proportional to FRO ancestry (figs. S10 and S11), a procedure that provides a sensitive test of Polynesian-specific admixture. Late prehistoric individuals from Pohnpei closely track the baseline, providing no evidence of Polynesian admixture. One present-day Micronesian (Jk2812) deviates from the line (maximum $|Z| = 3.3$) (table S21). We do not have a record of the island from which this individual came, so characterization of the Polynesian impact on Micronesia will require further sampling.

A working model for Micronesian population history

We started with a model previously used to study southwest Pacific lineages (*8*, *9*) and then added lineages and admixture events, testing alternative models for fit (Fig. 4C and figs. S12 to S15). With so many populations, the space of possible admixture graph topologies is vast, and the topology we show is unlikely to be a the only fit to the *f*-statistics. Nevertheless, identifying an admixture graph model is useful to demonstrate that all the features described in our analysis of individual *f*-statistics can jointly fit the data. We confirmed key inferences about admixture proportions and closest phylogenetic relatives of analyzed groups using *qpWave* and *qpAdm* (tables S22 to S25), which does not require making specific assumptions about deep phylogenetic relationships and allows us to test whether there are any groups that harbor genetic drift that is not present in the populations used as proxies for their ancestry (*20*). Finally, we used admixture linkage disequilibrium to estimate the ages of some detected admixture events with the software DATES (Fig. 4D and table S27).

(i) The Mariana Islands: Distinctive FRO ancestry without Papuan admixture. The Unai individuals from Guam whose radiocarbon dates range from 2800 to 2200 BP derive from the $\text{FRO}_{\text{Marianas}}$ lineage (M1) and have homogeneous ancestry. Later Latte individuals from Guam and Saipan after 700 BP derive ~85% of their ancestry from the same source (fig. S13A), with substantial continuity also confirmed by their harboring the same mitochondrial haplogroups E1 and E2 that are seen in the Unai period. The Latte individuals also derived ~15% ancestry from a previously unidentified $\text{FRO}_{\text{Palau}}$ lineage (M2), which we estimate mixed with $\text{FRO}_{\text{Marianas}}$ 45 to 50 generations before the Latte individuals lived (2400 to 1700 BP, assuming 28 years per generation). The admixture date shows that this migration and mixture process cannot be invoked to explain the origin of the Latte archaeological phenomenon in the Mariana Islands, which began much later at ~1100 BP.

The modern CHamoru from Guam are admixed with European (~19%) and Native American (~9%) ancestry (Fig. 4B), plausibly associated with Spanish colonial activities from the mid-16th century onward (*29*). Their remaining ancestry is entirely FRO. Although our analyses of modern CHamoru did not allow us to unambiguously determine their FRO source, they show a greater genetic affinity to $\text{FRO}_{\text{Marianas}}$ than to $\text{FRO}_{\text{SouthwestPacific}}$ (Fig. 3B), and their mitochondrial haplogroups E1 and E2 are also found in the Unai and Latte individuals, suggesting that they derived much of their East Asian–associated ancestry from earlier groups in Guam.

(ii) Palau: Mixture of $\text{FRO}_{\text{Palau}}$ and $\text{Papuan}_{\text{NewGuinea}}$ ancestry. Present-day Palauans are inferred to have ~62% $\text{FRO}_{\text{Palau}}$ ancestry (M2) from the same lineage that admixed in a smaller proportion into the Latte individuals (fig. S13A) and ~38% $\text{Papuan}_{\text{NewGuinea}}$ ancestry (M4). We estimate the date of $\text{FRO}_{\text{Palau}}$-$\text{Papuan}_{\text{NewGuinea}}$ admixture to be ~2500 to 2200 BP, suggesting the possibility of Papuan migration into this region by this time.

(iii) Central Micronesia: Mixture of $\text{FRO}_{\text{SouthwestPacific}}$ and $\text{Papuan}_{\text{NewGuinea}}$. We infer genetic homogeneity in central Micronesia over space and time, with Pohnpei and Chuuk having similar proportions of ~73% $\text{FRO}_{\text{SouthwestPacific}}$ (M3) and ~27% $\text{Papuan}_{\text{NewGuinea}}$ ancestry (M4) and forming a clade with the 11 individuals from prehistoric Pohnpei (Fig. 4B). $\text{FRO}_{\text{SouthwestPacific}}$ is a better single-source proxy for the primary First Remote Oceanian ancestry in Central Micronesia than $\text{FRO}_{\text{Marianas}}$, but an entirely $\text{FRO}_{\text{SouthwestPacific}}$ source fails in *qpAdm* when Unai and Latte are included as outgroups, suggesting that both $\text{FRO}_{\text{SouthwestPacific}}$ and $\text{FRO}_{\text{Marianas}}$ contributed. These findings also illuminate the origins of Nuclear Micronesian languages. Central Micronesians lack the Papuan ancestry that is predominant in the Solomon Islands, providing evidence against one of the three main candidate geographic regions (*24*). They also lack the $\text{Papuan}_{\text{NewBritain}}$ signature that was ubiquitous in Vanuatu by the time of the peopling of Central Micronesia, providing evidence against another candidate region. Instead, *qpAdm* shows that the people of Manus are a better proximate source for the $\text{Papuan}_{\text{NewGuinea}}$ ancestry than those of mainland New Guinea (table S24), increasing the likelihood of the third candidate—the Admiralty Islands—as the source for these languages and for the stream of migration that brought them. This should not be interpreted as implying that people specifically from Manus Island were the true source, but rather that the source was probably a genetically similar population from the Admiralty Islands or a coastal region along the northern fringe of mainland New Guinea.

We infer dates of $\text{FRO}_{\text{SouthwestPacific}}$-$\text{Papuan}_{\text{NewGuinea}}$ mixture in Chuuk and Pohnpei of 2100 to 1800 BP, showing that these lineages came into contact at least by the time of the peopling of Central Micronesia around 2000 BP and raising the possibility that the M3 and M4 lineage expansions into Central Micronesia came as part of an already mixed stream of people speaking early Nuclear Micronesian. An alternative, however, would accommodate a different perspective on the origins of Nuclear Micronesian languages, allowing M3 to have come from a $\text{FRO}_{\text{SouthwestPacific}}$ group that spoke a Southeast Solomonic language (*30*), to be joined later by an M4 Papuan-Admiralties group that did not displace already established Nuclear Micronesian languages. Such a scenario of language continuity despite population replacement would parallel the situation posited for Vanuatu (*8*, *9*). We do not yet have data from Yap but, given that Yapese is an earlier branching Proto-Oceanic language, we hypothesize that the indigenous Yap islanders might derive from a different mixture of source populations than other Central Micronesians.

Matrilocality in early Pacific islanders

We observed a notable degree of mitochondrial DNA differentiation between the $\text{FRO}_{\text{Marianas}}$ and the $\text{FRO}_{\text{SouthwestPacific}}$ lineages. All of the Unai individuals with mitochondrial haplogroup determinations and without evidence of high contamination carried haplogroups E1 and E2 (table S2), whereas all of the Lapita individuals had haplogroup B4 (*7*–*10*). All three haplogroups were found in Iron Age Taiwanese (*5*, *6*), consistent with

the finding that the Iron Age Taiwanese were relatively undrifted descendants of a population that was also ancestral to the Unai and Lapita individuals. Such a high level of mitochondrial differentiation is surprising given the intermediate degree of autosomal differentiation as measured by F_{ST}, a standard statistic measuring population genetic differentiation, which is 0.083 between the Unai and Lapita groups. This raises the possibility of greater genetic drift on the maternal than paternal line during the early divergence and radiation of FRO lineages.

We carried out simulations to determine the probability that completely different mitochondrial macrohaplogroups spread over the two populations since they diverged, under the null assumption that males and females had the same demographic behavior and given the observed genetic drift on the autosomes (fig. S16). This null hypothesis is rejected (P = 0.0014, Fisher's exact test) (*31*). The P values are not sensitive to assumptions about the split time of the $FRO_{Marianas}$ and the $FRO_{SouthwestPacific}$ lineages (table S28). These patterns are qualitatively opposite to those in Neolithic and Bronze Age Europe, where patrilocal patterns of greater female than male mobility among households have been inferred by analyzing ancient DNA data (*32*, *33*). Matrilocality in early Remote Oceanians has been hypothesized based on genetic and ethnographic studies of present-day communities, many of which have matrilocal practices in which women tend to raise their children in the same households in which they grew up (*34*, *35*). Our results provide direct evidence for the practice of matrilocality among FRO populations.

These findings concerning matrilocality among the ancestors of Lapita and Unai individuals with little if any Papuan ancestry are not related to previous evidence of sex-biased admixture between Papuan and FRO ancestry in some Pacific populations (*7*). However, a new finding of this study does concern sex-biased mixture. Specifically, we find that the Papuan ancestry in Palau and Central Micronesia was primarily derived from male ancestors, based on significantly more Papuan ancestry on the autosomes than on the X chromosome ($|Z|$> 2.2 to 3.3) (Fig. 4E and table S29) (*7*). This is notable because each of the three cases of FRO-Papuan admixture that are now documented (Palau, Central Micronesia, and southwest Pacific and Polynesia) involved a different pair of Papuan and FRO groups. These events must have been independent, and yet all share the feature of Papuan ancestry being transmitted primarily by male ancestors.

Family structure and population size during the Latte period

We measured runs of homozygosity (ROH) that were longer than 4 centimorgans (cM) for 113 Latte individuals with high-enough-quality data to allow such analyses (table S30). Only two had single stretches of ROH longer than 50 cM, indicating that close-kin unions were avoided in Latte people. Nine individuals from Guam and nine from Saipan had at least one ROH longer than 20 cM, suggesting that mating pairs of close relatives such as second or third cousins on both islands were relatively common. Shorter ROH signals (>4 cM) were also abundant, implying a limited pool of reproductive partners in every generation. We estimated the size of the population from which the Latte individuals in Guam and Saipan were drawing their reproductive partners to be 315 to 356 individuals in Guam and 361 to 424 individuals in Saipan (table S32).

We further analyzed long shared DNA segments [identical by descent (IBD) blocks] between the X chromosomes of male individuals (one from Guam and the other from Saipan). We identified 149 pairs of individuals who shared IBD segments longer than 8 cM (table S31). This puts an upper bound on N_e, the size of the mating population in the combined Mariana Islands, of 1203 to 1712 (95% confidence interval) (table S32). If there were restricted migration between islands, or if there were temporal variation in the dates of the individuals we compared, these numbers would be overestimates. This implies a long-term small population size or strong founder event in Latte history.

We identified 122 pairs of closely related Latte individuals (up to third-degree relatives) (fig. S17 and table S33). Eighty of 125 Latte individuals that were studied had one or several close relatives.

Discussion

A notable finding of this study is that the phenomenon of primarily male Papuan migrants mixing with previously resident FRO populations ~2500 to 2000 BP occurred at least three times, because the pairs of mixing sources were different in three regions (Fig. 4D). One of these migration and mixture processes occurred at an average date of ~2500 to 2200 BP, with $Papuan_{NewGuinea}$-FRO_{Palau} mixture forming modern Palauans. A second occurred ~2300 to 1600 BP, with $Papuan_{NewGuinea}$-$FRO_{SouthwestPacific}$ mixture forming ancient and modern Central Micronesians. A third occurred ~2300 to 1500 BP, with $Papuan_{NewBritain}$-$FRO_{SouthwestPacific}$ mixture forming the ancestry of ancient and modern people of the southwest Pacific and Polynesia (*7*). All three mixtures were sex asymmetric, with most of the Papuan ancestry deriving from males (Fig. 4C). Even in the Mariana Islands, where there is no evidence of Papuan mixture, the inferred FRO_{Palau}-$FRO_{Marianas}$ mixture date in Latte individuals is ~2400 to 1700 BP, providing a fourth example of migration and mixture in Remote Oceania occurring on average ~2500 to 2000 BP, well after the initial peopling events that involved entirely FRO groups.

A high-resolution ancient DNA time transect in Vanuatu has revealed the dynamics of this process in the southwest Pacific, where an initial $FRO_{SouthwestPacific}$ migration stream likely from New Britain changed into a primarily male $Papuan_{NewBritain}$ stream in the late Lapita period, likely deriving from the same source region and following previously established communication routes (*36*). Our results raise the possibility of similar processes in at least two other regions. The oldest pottery discovered in Pohnpei at ~2000 BP, which resembles that of late Lapita (*19*), provides an archaeological correlate for a spread of mixed $FRO_{SouthwestPacific}$-$Papuan_{NewGuinea}$ ancestry into this region. Parallel processes could have drawn $Papuan_{NewGuinea}$ ancestry into Palau and FRO_{Palau} ancestry into the Mariana Islands.

Our identification of the FRO_{Palau} lineage raises the possibility that the three FRO lineages correspond to the first-order three language splits in Malayo-Polynesia: $FRO_{Marianas}$ leading to the CHamoru language and associated with the Unai burials dated to ~2800 BP; $FRO_{SouthwestPacific}$ leading to CEMP languages and associated with the Lapita archaeological complex and burials dating to ~3000 BP in Vanuatu; and FRO_{Palau} bringing ancestral Palauan and plausibly the first ancestry type in Palau because mitochondrial DNA of 3000 to 1800 BP remains from Chelechol ra Orrak suggests East Asian ancestry (*37*).

The ordering of the FRO lineage splits is also important. The fact that the FRO_{Palau} lineage split first cannot be explained by the theory that there was a single First Remote Oceanian spread into the Mariana Islands (*28*, *38*), which then gave rise to the other lineages, because in this case, $FRO_{Marianas}$ would have branched first. The theory of a Mariana population being ancestral to all FRO lineages is further challenged by the mitochondrial DNA evidence. If this theory were correct, the most parsimonious expectation is for the haplotypes observed in the Unai individuals from Guam at 2800 to 2200 BP (E1 and E2) also to be observed in the Lapita individuals at 3000 to 2500 BP. However, only mitochondrial haplotype B4a1a1 (the "Polynesian motif") is observed. Therefore, our results point to a scenario in which three First Remote Oceanian lineages branched from a trunk of MP speakers in Island Southeast Asia, with at least three independent streams of migrations into Remote Oceania.

Since colonial times, Pacific peoples have been divided into "Melanesians," "Polynesians," and "Micronesians," driven by theories

of shared origins (*39*). However, our results show that people in Micronesia have a diversity of ancestral origins even within the same geographic region, implying that the term "Micronesian" should be used as a geographic label without implying a specific biological profile.

REFERENCES AND NOTES

1. J. F. O'Connell, J. Allen, *J. Archaeol. Sci.* **56**, 73–84 (2015).
2. S. Wickler, M. Spriggs, *Antiquity* **62**, 703–706 (1988).
3. T. M. Rieth, J. S. Athens, *J. Isl. Coast. Archaeol.* **14**, 5–16 (2019).
4. S. Bedford, M. Spriggs, Eds., *Debating Lapita: Distribution, Chronology, Society and Subsistence*, vol. 52 of *Terra Australis* (ANU Press, 2019).
5. M. A. Yang *et al*., *Science* **369**, 282–288 (2020).
6. C.-C. Wang *et al*., *Nature* **591**, 413–419 (2021).
7. P. Skoglund *et al*., *Nature* **538**, 510–513 (2016).
8. M. Lipson *et al*., *Curr. Biol.* **28**, 1157–1165.e7 (2018).
9. M. Lipson *et al*., *Curr. Biol.* **30**, 4846–4856.e6 (2020).
10. C. Posth *et al*., *Nat. Ecol. Evol.* **2**, 731–740 (2018).
11. M. T. Carson, *First Settlement of Remote Oceania: Earliest Sites in the Mariana Islands*, vol. 1 of *Springer Briefs in Archaeology* (Springer, 2014).
12. F. Petchey, G. Clark, *J. Isl. Coast. Archaeol.* **5**, 236–252 (2010).
13. F. Petchey *et al*., *Quat. Geochronol.* **48**, 180–194 (2018).
14. F. Petchey, G. Clark, *Radiocarbon* **63**, 905–913 (2021).
15. A. Spoehr, *Marianas Prehistory: Archaeological Survey and Excavations on Saipan, Tinian and Rota*, vol. 48 of *Fieldiana Anthropology* (Chicago Natural History Museum, 1957).
16. G. R. Clark, O. Winter, in *Debating Lapita: Distribution, Chronology, Society and Subsistence*, vol. 52 of *Terra Australis*, S. Bedford, M. Spriggs, Eds. (ANU Press, 2019), pp. 37–59.
17. T. Rieth, E. E. Cochrane, in *The Oxford Handbook of Prehistoric Oceania*, E. E. Cochrane, T. L. Hunt, Eds. (Oxford Univ. Press, 2018), pp. 133–161.
18. R. Shutler Jr., in *The Pacific from 5000 to 2000 BP (Colonsation and Tranformations)*, J. C. Galipaud, I. Lilley, Eds. (IRD Editions, 1999), pp. 522–529.
19. J. S. Athens, *Micronesica* **2**, 17–32 (1990).
20. See supplementary materials.
21. R. D. Gray, F. M. Jordan, *Nature* **405**, 1052–1055 (2000).
22. R. D. Gray, A. J. Drummond, S. J. Greenhill, *Science* **323**, 479–483 (2009).
23. A. D. Smith, *Ocean. Linguist.* **56**, 435–490 (2017).
24. J. J. Song, *Macrolinguistics* **3**, 26–66 (2009).
25. M. Ross, in *Reconstruction, Classification, Description — Festschrift in Honor of Isidore Dyen*, vol. 3 of *Abera Network Asia-Pacific*, B. Nothofer, Ed. (Abera, 1996) pp. 121–166.
26. P. V. Kirch, *J. Pac. Hist.* **19**, 224–238 (1984).
27. P. V. Kirch, *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact* (Univ. California Press, 2017).
28. I. Pugach *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022112118 (2021).
29. W. L. Schurz, *The Manila Galleon*, vol. 40 of *Historical Conservation Society* (E. P. Dutton & Company, 1939).
30. R. Blust, *Ocean. Linguist.* **49**, 559–567 (2010).
31. K.-I. Sudo, *Senri Ethnol. Stud.* **17**, 203–230 (1984).
32. A. Mittnik *et al*., *Science* **366**, 731–734 (2019).
33. C. Fowler *et al*., *Nature* **601**, 584–587 (2022).
34. P. Hage, J. Marck, *Curr. Anthropol.* **44**, S121–S127 (2003).
35. F. M. Jordan, R. D. Gray, S. J. Greenhill, R. Mace, *Proc. Biol. Sci.* **276**, 1957–1964 (2009).
36. M. Spriggs, D. Reich, *World Archaeol.* **51**, 620–639 (2019).
37. J. H. Stone, "The bioarchaeology of initial human settlement in Palau, Western Micronesia," thesis, University of Oregon (2020).
38. M. T. Carson, H. Hung, G. Summerhayes, P. Bellwood, *J. Island Coast. Archaeol.* **8**, 17–36 (2013).
39. J. S. C. Dumont D'Urville, *Bull. la Société Géographie Paris* **105**, 1–21 (1832).
40. Y.-C. Liu *et al*., Ancient DNA Reveals Five Streams of Migration into Micronesia and Matrilocality in Early Pacific Seafarers Harvard Dataverse (2022); https://doi.org/10.7910/DVN/63QFEC.

ACKNOWLEDGMENTS

We acknowledge the people past and present who were the source of the samples we analyzed. We are grateful to CHamoru community members to whom we presented results on 15 October 2020 and to the Pohnpei Historic Preservation Office for review and approval of this study on 6 September 2020; we incorporated feedback from both these engagements into the final manuscript. For help in the collections of modern DNA, we thank E. Pretrick, M. Kumangai, and A. Loerzel. We also thank R. Blust, R. Chong-Cruz, P. Flegontov, E. Harney, P. Iohn, R. Lemuel, I. Lazaridis, R. Maier, R. Palomo, T. Parks, H. Ringbauer, L. T. Souder, and L. M. Young for critical comments. **Funding:** This work was supported by grants from the National Institutes of Health (GM100233 and HG012287), the John Templeton Foundation (grant 61220), and the Allen Discovery Center program, a Paul G. Allen Frontiers Group–advised program of the Paul G. Allen Family Foundation. D.R. is an investigator of the Howard Hughes Medical Institute. T.G.S. was supported by the National Geographic Society and the University of Pennsylvania. R.H.-A. and J.E. were supported by the Guam Preservation Trust and the National Geographic Society. **Author contributions:** R.H.-A., J.E., R.P., and D.R. conceived the project. N.P., M.L., R.P., and D.R. supervised the study. R.H.-A., O.C., J.E., M.P., J.S.A., R.M.I.-Q., and M.T.D. sampled prehistoric specimens. R.H.-A., J.E., M.P., J.S.A., R.M.I.-Q., M.T.D., and M.S. assembled archaeological and anthropological information. F.C., A.I., K.K., J.K., T.G.S., A.V.S.H., A.J.M., K.R., K.A., C.D.Q.-C., C.D.B., A.M.-E., and M.V. gathered data from present-day populations. N.R., R.B., M.Ma., K.M., M.Mi., N.B., J.O., N.A., K.S., A.M.L., F.Z., K.C., and T.G.S. conducted laboratory and/or data processing work. B.J.C. and D.J.K. performed radiocarbon dating analyses. Y.-C.L., N.P., M.L., and D.R. analyzed the data. Y.-C.L. and D.R. wrote the manuscript. **Competing interests:** C.D.B. is founder and CEO of Galatea Bio. The authors declare no other competing interests. **Data and materials availability:** Data for the prehistoric individuals are fully publicly available and have been deposited in the European Nucleotide Archive (project accession no. PRJEB51180). The informed consents for the newly genotyped present-day individuals from Guam, Palau, Chuuk, and Pohnpei are not consistent with unmediated public posting of genomic data. Researchers who wish to analyze these deidentified data can access them through the Harvard Dataverse repository (*40*). Data may be downloaded after registering for a Harvard Dataverse user account, providing an email address and institutional or professional affiliation, and submitting an affirmation of the following statements: (i) I will not distribute the data outside my collaboration; (ii) I will not post the data publicly; (iii) I will make no attempt to connect the genetic data to personal identifiers for the samples; (iv) I will use the data only for studies of population history; and (v) I will not use the data for commercial purposes.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm6536
Materials and Methods
Supplementary Text
Figs. S1 to S23
Tables S1 to S38
References (*41–127*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 1 October 2021; accepted 18 May 2022
10.1126/science.abm6536

NEUROSCIENCE

Microglia-mediated degradation of perineuronal nets promotes pain

Shannon Tansley[1,2], Ning Gu[1], Alba Ureña Guzmán[1], Weihua Cai[1], Calvin Wong[1], Kevin C. Lister[1], Einer Muñoz-Pino[3], Noosha Yousefpour[4], R. Brian Roome[5], Jordyn Heal[1], Neil Wu[1], Annie Castonguay[3], Graham Lean[6], Elizabeth M. Muir[7], Artur Kania[5,8], Masha Prager-Khoutorsky[6], Ji Zhang[9,10,11], Christos G. Gkogkas[12], James W. Fawcett[13,14], Luda Diatchenko[1,10,11], Alfredo Ribeiro-da-Silva[4,8,11], Yves De Koninck[3,4,11], Jeffrey S. Mogil[1,2,11]*, Arkady Khoutorsky[1,10,11]*

Activation of microglia in the spinal cord dorsal horn after peripheral nerve injury contributes to the development of pain hypersensitivity. How activated microglia selectively enhance the activity of spinal nociceptive circuits is not well understood. We discovered that after peripheral nerve injury, microglia degrade extracellular matrix structures, perineuronal nets (PNNs), in lamina I of the spinal cord dorsal horn. Lamina I PNNs selectively enwrap spinoparabrachial projection neurons, which integrate nociceptive information in the spinal cord and convey it to supraspinal brain regions to induce pain sensation. Degradation of PNNs by microglia enhances the activity of projection neurons and induces pain-related behaviors. Thus, nerve injury–induced degradation of PNNs is a mechanism by which microglia selectively augment the output of spinal nociceptive circuits and cause pain hypersensitivity.

Peripheral nerve injury leads to long-lasting pain hypersensitivity. Damaged primary afferents release chemokines, signaling molecules, and proteases to activate spinal cord microglia, which in turn enhance the excitability of spinal nociceptive circuits (*1*, *2*). Microglia release an array of bioactive substances that bind to cell surface receptors to increase neuronal activity through modulation of intracellular processes (*1*, *2*). How these substances specifically sensitize pain-processing circuits without affecting other modalities is not well understood. Studies of mechanisms by which microglia affect neuronal functions in the spinal cord have focused on intracellular mechanisms of action rather than modulation of the extracellular matrix. The extracellular matrix in the central nervous system (CNS) not only provides structural support but is also involved in the regulation of neuronal excitability and synaptic plasticity (*3*, *4*). Its role in the regulation of spinal pain circuits, however, remains ill defined. Perineuronal nets (PNNs) are the most prominent extracellular matrix structures in the CNS and are composed of a proteoglycan core protein decorated by chondroitin sulfate sugar chains (*3*). In the neocortex, PNNs preferentially enwrap neuronal soma and proximal dendrites of fast-spiking parvalbumin-positive inhibitory interneurons (*5*), modulating their potential for neuroplasticity by regulating synaptic inputs (*6*) and intrinsic excitability (*7*). We investigated whether PNNs are found around and affect the activity of spinal cord neurons involved in the processing of nociceptive information.

[1]Department of Anesthesia, McGill University, Montreal, QC, Canada. [2]Department of Psychology, Faculty of Science, McGill University, Montreal, QC, Canada. [3]Department of Psychiatry and Neuroscience, CERVO Brain Research Centre, Université Laval, Québec, QC, Canada. [4]Department of Pharmacology and Therapeutics, McGill University, Montreal, QC, Canada. [5]Institut de Recherches Cliniques de Montreal (IRCM), Montreal, QC, Canada. [6]Department of Physiology, McGill University, Montreal, QC, Canada. [7]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. [8]Department of Anatomy and Cell Biology, McGill University, Montreal, QC, Canada. [9]Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada. [10]Faculty of Dental Medicine and Oral Health Sciences, Montreal, QC, Canada. [11]Alan Edwards Centre for Research on Pain, McGill University, Montreal, QC, Canada. [12]Biomedical Research Institute, Foundation for Research and Technology-Hellas, University Campus, 45110 Ioannina, Greece. [13]John van Geest Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [14]Centre for Reconstructive Neuroscience, Institute for Experimental Medicine CAS, Prague, Czech Republic.
***Corresponding author. Email: jeffrey.mogil@mcgill.ca (J.S.M.); arkady.khoutorsky@mcgill.ca (A.Kh.)**

PNNs in lamina I surround projection neurons

PNNs can be identified by staining with *Wisteria* floribunda agglutinin (WFA), which selectively binds the glycosaminoglycan (GAG) sugar side chains of PNN glycoproteins. Additionally, PNNs can be visualized by labeling their core protein component, aggrecan (*3*, *8*, *9*). Immunostaining of spinal cord sections for aggrecan showed that PNNs surround neurons with large-diameter soma in lamina I of the dorsal horn that exhibit a medio-lateral orientation (Fig. 1A). High-resolution confocal Airyscan imaging revealed presynaptic inhibitory terminals located within PNN holes (Fig. 1, B and C). Retrograde tracing with Fluoro-Gold (FG) injected into the lateral parabrachial (LPb) nucleus showed that PNNs in lamina I are present selectively around spinoparabrachial projection neurons (*10*, *11*) and not found around other cell types (Fig. 1, D and E). Lamina I PNN-positive (PNN^+) projection neurons have significantly larger soma than those of PNN^--negative (PNN^-) projection neurons, or other neuronal cell types (volumetric values are provided in Fig. 1F, and soma diameters are provided in fig. S1A), and are $NK1R^+$ and $Phox2a^-$ (fig. S1, B to E). PNNs are absent in lamina II and are sparsely found throughout lamina III of the dorsal horn (Fig. 1A). In laminae IV and V, PNNs are found around both inhibitory ($Pax2^+$) and excitatory ($Pax2^-$ and $NeuN^+$) neurons (Fig. 1A). Exposure of mouse hind paw to mechanical (25-g binder clip for 30 s) (Fig. 1, G and I) or thermal (55°C water bath for 30 s) (Fig. 1, H and I) noxious stimuli induced Fos expression in PNN^+ lamina I projection neurons (thermal stimuli, in 70.6 ± 1.7% of neurons; mechanical stimuli, in 49.3 ± 5.2% of neurons), suggesting that these neurons are involved in the processing of pain-related information.

Modification of PNNs after peripheral nerve injury

To study whether PNNs in the spinal cord are modified after peripheral nerve injury, we subjected mice to the spared nerve injury (SNI) assay of neuropathic pain, which features pain behaviors at maximal levels by 2 to 3 days after injury (*12*, *13*). The intensity of WFA staining, which labels GAGs on PNNs (Fig. 1J), around lamina I projection neurons was significantly reduced 3 days after SNI (decrease of 76.3%) (Fig. 1, K and L), whereas aggrecan staining was not affected. WFA staining around lamina I projection neurons remained low at 7 and 14 days after nerve injury (Fig. 1L). PNNs in deeper laminae did not show a decrease in WFA staining (fig. S1, F and G).

PNN degradation is mediated by microglia

Three days after SNI, numerous microglia contained WFA immunoreactivity within their lysosomes, identified by means of labeling with antibody to CD68 (Fig. 2A). Microglia showing CD68 and WFA colocalization were present at day 3 after SNI (67.1% of all microglia in the dorsal horn) (Fig. 2, A and B; sex-disaggregated analysis is available in fig. S1, H, I, and J) but not at later time points (days 7 and 14). To study the causal role of microglia in the degradation of PNNs after nerve injury, we depleted microglia using Cre-inducible mice expressing diphtheria toxin receptor (DTR) (*14*) selectively in microglia (iDTR;$TMEM119^{CreERT2}$) (the treatment regimen with tamoxifen to induce Cre expression and diphtheria toxin to ablate microglia is provided in Fig. 2C, and the effect on microglia is provided in Fig. 2D) (*15*). Peripheral nerve injury in microglia-depleted mice did not evoke mechanical hypersensitivity (Fig. 2E: sex-disaggregated analysis is available in fig. S1K) and did not induce degradation of PNNs because no significant reduction in WFA signal around lamina I projection neurons was observed at day 3 after SNI (Fig. 2, F and G).

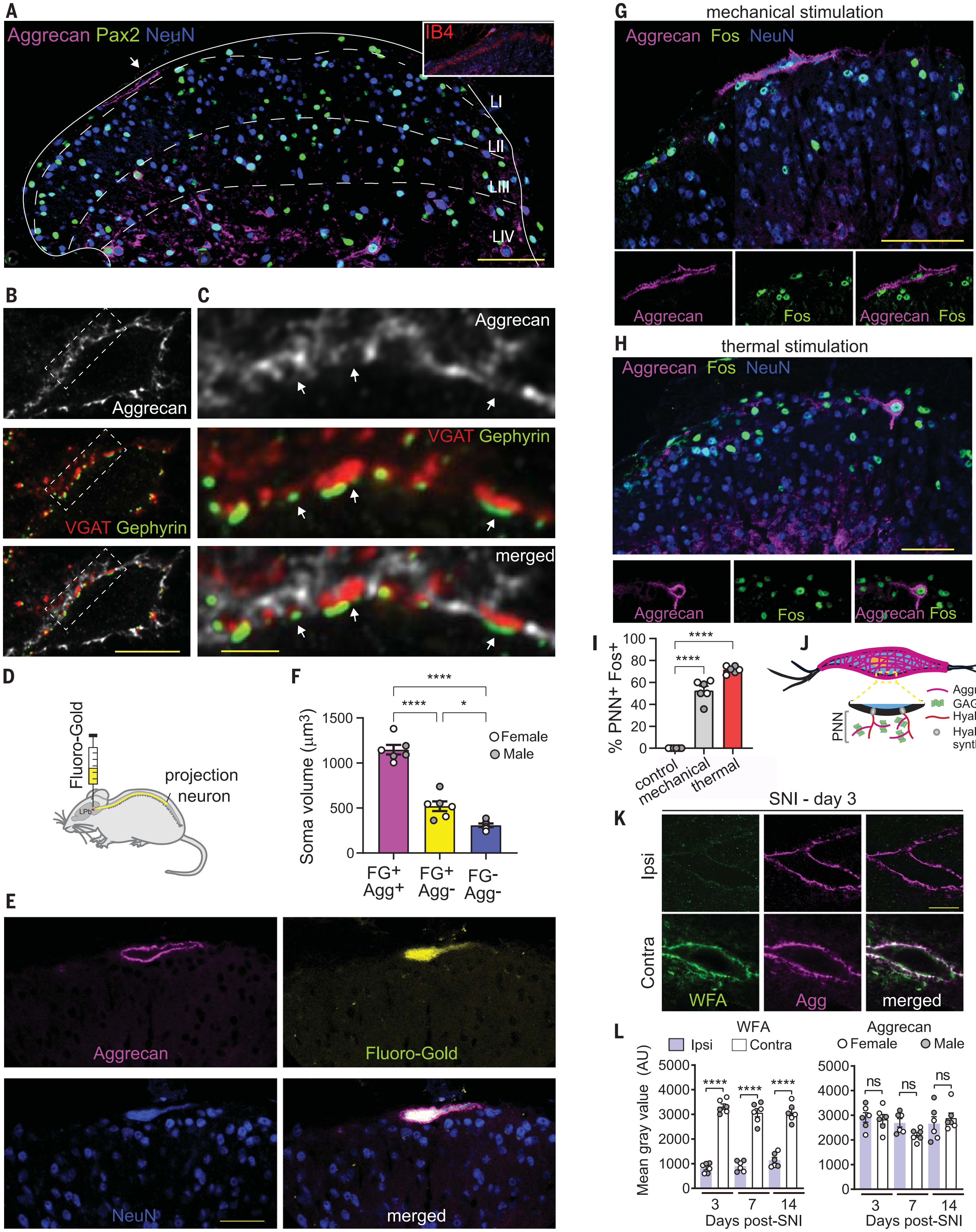


Fig. 1. PNNs are found around lamina I projection neurons and are modified after peripheral nerve injury. (**A**) Immunostaining for a marker of inhibitory neurons (Pax2), all neurons (NeuN), and aggrecan. PNNs are present around large-diameter neurons in lamina I (white arrow). Scale bar, 100 μm. (**B**) Low-magnification and (**C**) high-magnification Airyscan images. VGAT labels presynaptic inhibitory terminals; gephyrin labels postsynaptic inhibitory compartment. Scale bars, (B) 10 μm; (C) 2 μm. (**D**) Schematic illustration showing retrograde labeling of spinoparabrachial projection neurons in a mouse injected

with FG into the LPb nucleus. (**E**) Immunostaining against aggrecan shows that lamina I projection neurons are retrogradely labeled by FG. (**F**) Lamina I projection neurons surrounded by PNNs (Agg^+) have larger soma volume than those of PNN-negative projection neurons or other cell types [FG^+ Agg^+ versus FG^+ Agg^-, $q(15) = 13.8$, $P < 0.0001$; FG^+ Agg^+ versus FG^- Agg^-, $q(15) = 18.5$, $P < 0.0001$; $n = 6$ mice per group, one-way analysis of variance (ANOVA) followed by Tukey's post-hoc comparison]. (**G** and **H**) Images of Fos expression after (G) mechanical and (H) thermal stimuli. Scale bar, 100 μm. (**I**) Quantification of the percentage of lamina I PNN^+ projection neurons showing Fos immunoreactivity 60 min after the evoking mechanical and thermal stimuli. Each data point represents one animal ($n = 6$ per condition). (**J**) Schematic showing different components of PNNs. (**K**) Staining with WFA for GAG shows elimination of GAGs from PNNs at day 3 after SNI. Scale bar, 10 μm. (**L**) Quantification of WFA (left) and aggrecan (right) signal around lamina I projection neurons at day 3, day 7, and day 14 after SNI shows reduced WFA signal but no change in aggrecan levels [WFA, day 3 Ipsi versus Contra, $q(30) = 20.5$, $P < 0.0001$; day 7 Ipsi versus Contra, $q(30) = 17.8$, $P < 0.0001$; day 14 Ipsi versus Contra, $q(30) = 15.93$, $P < 0.0001$; one-way ANOVA followed by Tukey's post-hoc comparison, $n = 6$ mice per condition]. All data are presented as mean ± SEM. *$P < 0.05$; ****$P < 0.0001$; ns, not significant.

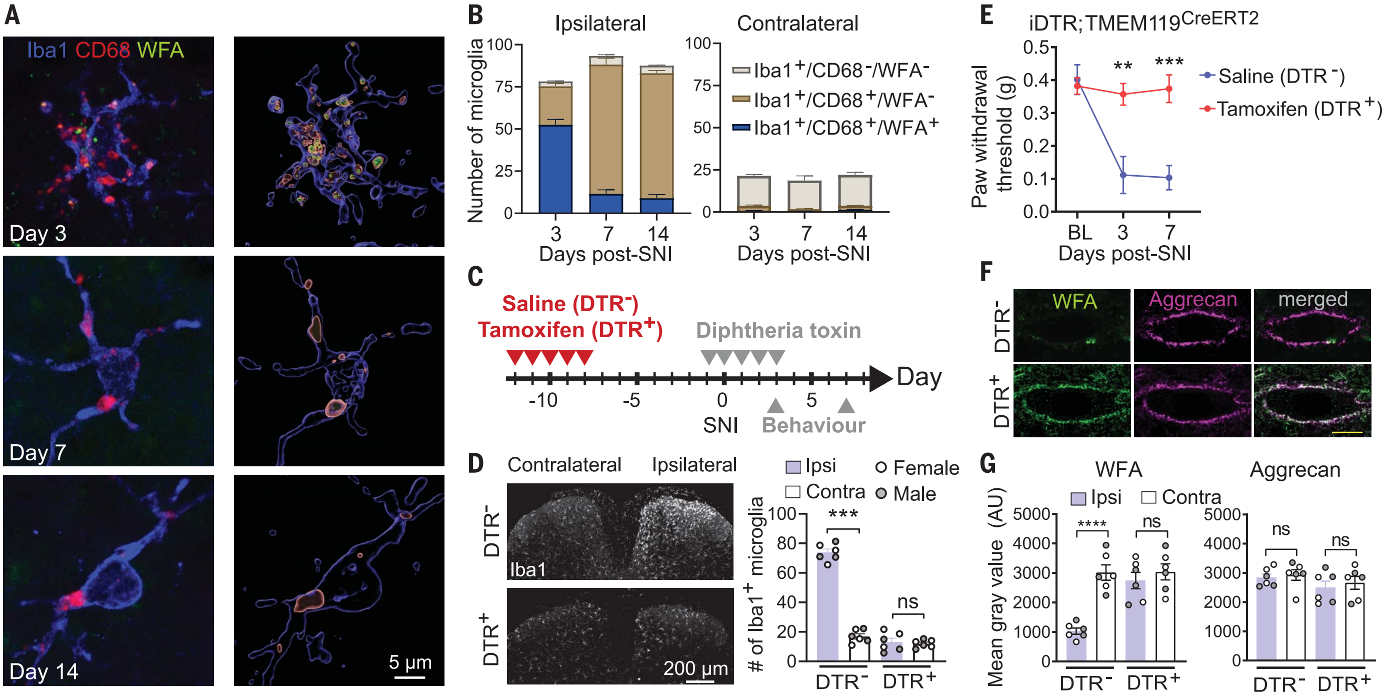


Fig. 2. Microglia mediate the degradation of PNNs around lamina I projection neurons. (**A**) (Left) Microglia (Iba1), lysosomes (CD68), and GAG (WFA) at day 3, day 7, and day 14 after SNI in the dorsal horn spinal cord. (Right) Reconstruction with Imaris software. (**B**) Quantification of the number of microglia per section with and without CD68 and WFA in the dorsal horn ($n = 6$ mice per condition). (**C**) Protocol for administration of tamoxifen and diphtheria toxin (DT) in iDTR;TMEM119^{CreERT2} mice. (**D**) Images showing $Iba1^+$ microglia in saline-treated (DTR^-, control) and tamoxifen-treated (DTR^+, microglia-depleted) iDTR;TMEM119^{CreERT2} mice (both groups received DT). Quantification on the right shows the number of microglia in both groups ($n = 6$ mice per group, one-way ANOVA followed by Tukey's post-hoc comparison). (**E**) SNI in microglia-depleted mice does not lead to the development of mechanical hypersensitivity as assessed with von Frey filaments [withdrawal threshold: day 3, DTR^- versus DTR^+, $t(66) = 4.241$, $P = 0.001$; day 7, DTR^- versus DTR^+, $t(66) = 4.7$, $P = 0.0002$; $n = 12$ mice per group, two-way ANOVA followed by Sidak's post-hoc comparison]. Images (**F**) and quantification (**G**) showing elimination of GAGs from PNNs in control mice but not in microglia-depleted mice on day 3 after SNI [WFA, DTR^-, Ipsi versus Contra, $q(20) = 8.3$, $P < 0.0001$; DTR^+, Ipsi versus Contra, $q(20) = 1,22$, $P = 0.82$; $n = 6$ mice per condition, one-way ANOVA followed by Tukey's post-hoc comparison]. Scale bar, 10 μm. All data are presented as mean ± SEM. **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$; ns, not significant.

To further corroborate the role of microglia in the modification of PNNs, we used mice lacking CX3C chemokine receptor 1 (CX3CR1) (*16*), which is involved in microglia activation and stimulation of their phagocytic activity (*17*–*19*). After peripheral nerve injury, *Cx3cr1*$^{-/-}$ mice showed microgliosis comparable with that of control animals (fig. S2, A and B) but exhibited a significantly reduced number of microglial lysosomes (identified with CD68 immunostaining) (fig. S2, C and D), indicating reduced phagocytic activity. *Cx3cr1*$^{-/-}$ mice also did not develop pain hypersensitivity (fig. S2E), which is consistent with previous reports (*20*). Nerve injury in *Cx3cr1*$^{-/-}$ animals did not trigger PNN degradation around lamina I projection neurons (fig. S2, F and G), similarly to that in microglia-depleted mice, and significantly reduced WFA accumulation was detected in microglia at day 3 after SNI (fig. S2, C, H, and I).

Disruption of PNNs around projection neurons promotes pain-related behavior

To study whether the removal of PNNs around projection neurons can induce pain, we deleted aggrecan in projection neurons by injecting *Acan*$^{fl/fl}$ mice (*8*) with adeno-associated virus-retro (rAAV2)–Cre into the LPb nucleus (Fig. 3A). rAAV2-Cre allows retrograde labeling and expression of Cre recombinase in neurons projecting to the LPb (*21*, *22*), including lamina I spinoparabrachial projection neurons (characterization of rAAV2-Cre is available in fig. S3). We validated the disruption of PNNs around lamina I projection neurons in rAAV2-Cre–injected *Acan*$^{fl/fl}$ mice (Fig. 3, B and C). The elimination of PNNs around spinoparabrachial projection neurons caused thermal hypersensitivity; latencies to paw-licking and

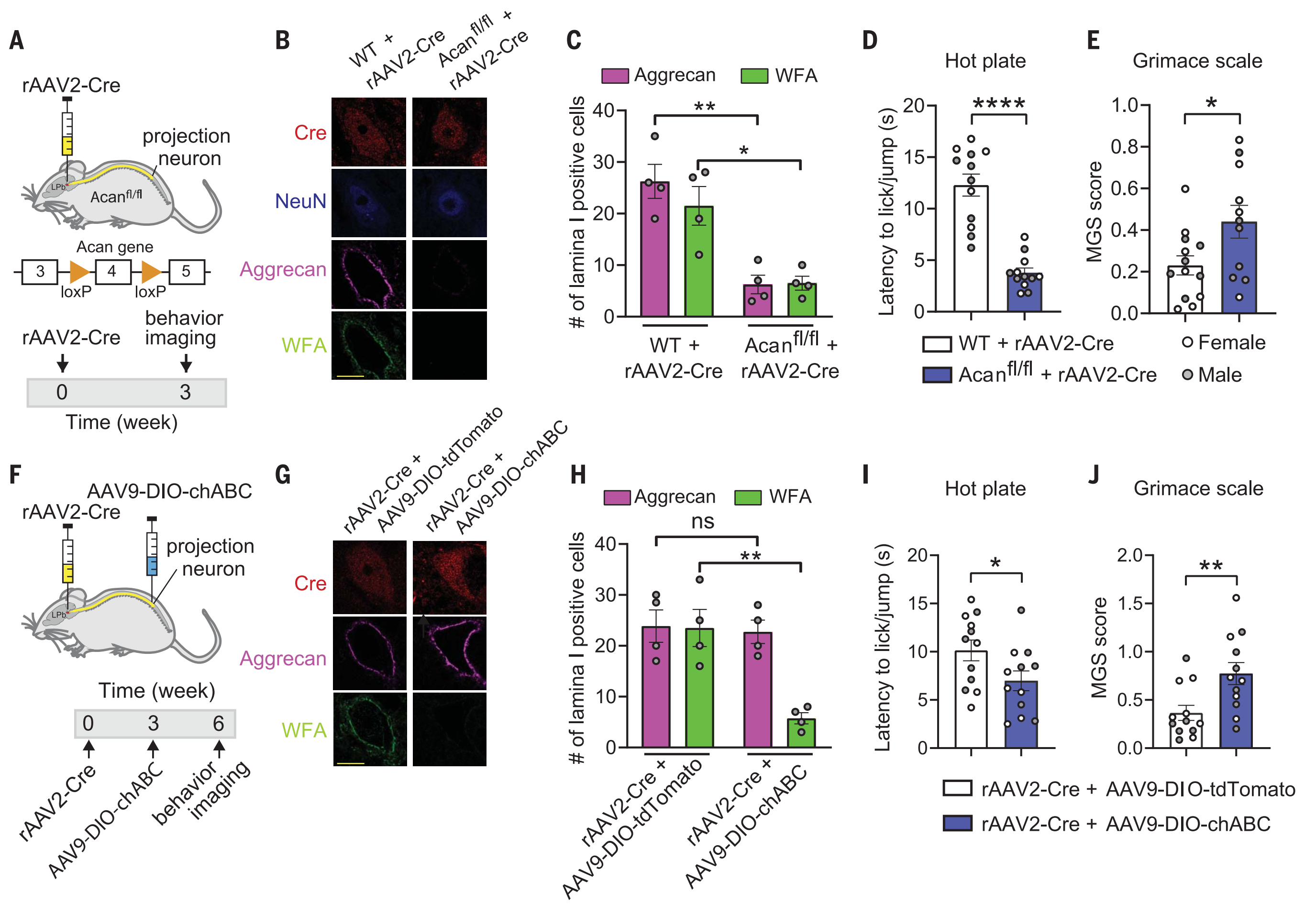


Fig. 3. Selective disruption of PNNs around lamina I projection neurons causes pain. (**A**) Schematic illustration showing the injection of rAAV2-Cre into the LPb nucleus of $Acan^{fl/fl}$ mice. (**B**) Lamina I spinoparabrachial projection neurons, identified by staining against Cre. (**C**) Quantification of aggrecan$^+$ and WFA$^+$ cells in lamina I of the lumbar spinal cord [aggrecan, wild type (WT) + rAAV2-Cre versus $Acan^{fl/fl}$ + rAAV2-Cre, $q(12) = 7.3$, $P = 0.0012$; WFA, WT + rAAV2-Cre versus $Acan^{fl/fl}$ + rAAV2-Cre, $q(12) = 5.5$, $P = 0.01$; $n = 4$ mice per condition, one-way ANOVA followed by Tukey's post-hoc comparison]. (**D** and **E**) Control and aggrecan-ablated mice were subjected to (D) hot plate and (E) Mouse Grimace Scale testing. Mice with ablation of aggrecan in lamina I spinoparabrachial projection neurons exhibit reduced time to licking and jumping in the hot plate test [(D) WT + rAAV2-Cre ($n = 12$ mice) versus $Acan^{fl/fl}$ + rAAV2-Cre ($n = 12$ mice), $P < 0.0001$; unpaired t test] and increased facial expressions of pain [(E) WT + rAAV2-Cre ($n = 13$ mice) versus $Acan^{fl/fl}$ + rAAV2-Cre ($n = 11$ mice), $P = 0.03$; unpaired Student's t test]. (**F**) chABC was expressed in lumbar spinoparabrachial projection neurons by means of coinjection of rAAV2-Cre into the LPb nucleus and Cre-dependent AAV9-DIO-chABC into the lumbar spinal cord. (**G**) Cre-expressing neurons show reduced WFA signal but no change in aggrecan levels. (**H**) Quantification of aggrecan$^+$ and WFA$^+$ cells in lamina I of the lumbar spinal cord [aggrecan: $q(12) = 0.411$, $P = 0.991$, WFA: $q(12) = 6.5$, $P = 0.003$; $n = 4$ mice per condition, one-way ANOVA followed by Tukey's post-hoc comparison]. (**I** and **J**) Mice expressing chABC in projection neurons show (I) reduced latency in the hot plate test [rAAV2-Cre + AAV9-DIO-tdTomato ($n = 12$ mice) versus rAAV2-Cre + AAV9-DIO-chABC ($n = 12$ mice); $P = 0.046$, unpaired Student's t test] and (J) increased spontaneous pain measured by the Mouse Grimace Scale [rAAV2-Cre + AAV9-DIO-tdTomato ($n = 12$ mice) versus rAAV2-Cre + AAV9-DIO-chABC ($n = 12$ mice); $P = 0.008$, unpaired Student's t test]. All data are presented as mean ± SEM. *$P < 0.05$; **$P < 0.01$; ****$P < 0.0001$; ns, not significant. Scale bar, 10 μm.

jumping in the hot plate test were significantly shortened in $Acan^{fl/fl}$ mice injected with rAAV2-Cre as compared with control animals (Fig. 3D). Removal of PNNs also elicited spontaneous pain as assessed with the Mouse Grimace Scale (Fig. 3E). Because peripheral nerve injury decreases GAGs on PNNs without affecting aggrecan (Fig. 1, L and M), and a limitation of the previous approach that all neurons projecting to the LPb nucleus are targeted, we next mimicked the effect of nerve injury on PNNs by selectively removing GAGs around lumbar spinal cord projection neurons by using adeno-associated virus (AAV)–expressing chondroitinase ABC (chABC) (*23*). ChABC specifically digests GAGs on PNNs and has been extensively used to study the role of PNNs in the nervous system (*6*, *24*, *25*). Expression of chABC in projection neurons in lumbar spinal cord was achieved by co-injecting two viral vectors: rAAV2-Cre into the LPb nucleus, and AAV-expressing chABC in a Cre-dependent manner (AAV9-DIO-chABC) into the lumbar spinal cord (Fig. 3F). This approach resulted in the removal of GAGs from lamina I PNNs without affecting aggrecan (Fig. 3, G and H). It did not change PNNs in deeper laminae (fig. S4A) or cause microglia activation (fig. S4B). Removal of GAGs from lumbar spinal cord projection neurons induced thermal hypersensitivity in the hot plate test (Fig. 3I) and evoked spontaneous pain in the Mouse Grimace Scale (Fig. 3J).

Removal of GAG disinhibits lamina I projection neurons

Previous studies in the cortex and hippocampus have revealed that PNNs regulate synaptic transmission (*25*) and affect intrinsic neuronal excitability (*7*). Our finding of PNN degradation promoting pain behavior (Fig. 3) prompted us to investigate whether removal of PNN GAGs

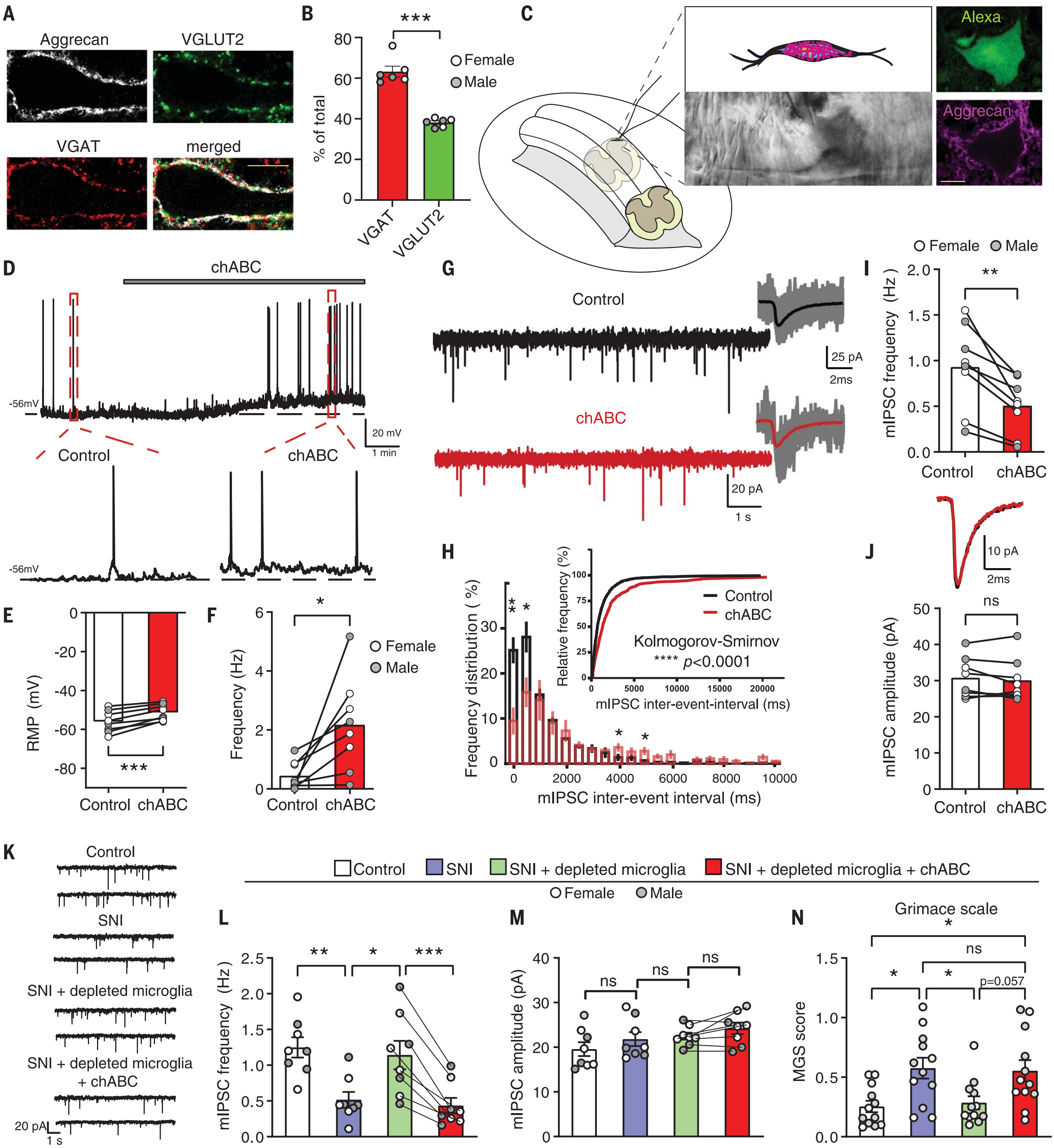


Fig. 4. Removal of PNNs increases projection neurons activity through disinhibition. (**A**) Immunostaining against presynaptic inhibitory (VGAT) and excitatory (VGLUT2) terminal markers around the cell body of PNN$^+$ lamina I projection neuron. (**B**) Quantification of (A) (P < 0.0001, n = 6 mice per condition, unpaired Student's t test). (**C**) Whole-cell recording in ex vivo spinal cord from large-diameter (>20 μm) lamina I neurons showing mediolateral orientation, which have PNNs. These cells were identified by use of oblique infrared illumination with LED. To verify the presence of PNNs around the patched cells, Alexa fluor 488 hydrazide was added to the patch electrode, followed by fixation and immunostaining for aggrecan. 82% of patched cells were positive for aggrecan (images on the right). (**D** to **F**) Current clamp recording shows that treatment with chABC induces membrane potential depolarization [(D) and (E), P = 0.00055, n = 8 cells from eight mice per group] and an increase in firing rate [(D) and (F), P = 0.022, n = 8 cells from eight mice per group, paired Student's t test]. (**G**) mIPSCs before and after treatment with chABC (0.2 U/ml for 10 min). (**H**) Frequency distribution histogram. (**I** and **J**) The frequency of mIPSCs is reduced after chABC treatment [(I) P = 0.003], but the mIPSC amplitude is not [(J) P = 0.69, n = 8 cells from eight mice per group,

paired Student's *t* test]. (**K** to **M**) mIPSC recorded from lamina I PNN^+ neurons in sham (control), SNI (3 to 5 days after surgery), SNI+depleted microglia (tamoxifen-treated iDTR;$TMEM119^{CreERT2}$ mice+DT), and SNI+depleted microglia+chABC (chABC applied onto ex vivo spinal cord) mice (*n* = 8 cells from eight mice per group, one-way ANOVA followed by Tukey's post-hoc comparison for the three first columns, paired Student's *t* test for the last two columns). (**N**) Mouse Grimace Scale testing on control, SNI (SNI+vehicle), SNI+depleted microglia (SNI+PLX5622+rAAV2-Cre+AAV9-DIO-tdTomato), and SNI+depleted microglia+chABC (SNI+PLX5622+rAAV2-Cre+AAV9-DIO-chABC, *n* = 12 animals per group, one-way ANOVA) (a detailed timeline is available in the supplementary materials, materials and methods). All data are presented as mean ± SEM. *P < 0.05; **P < 0.01, ***P < 0.001; ns, not significant. Scale bar, 10 μm.

increases the activity of lamina I projection neurons in the dorsal horn of the spinal cord.

First, we assessed the number of excitatory and inhibitory synaptic inputs onto PNN^+ projection neurons by quantifying inhibitory [vesicular γ-aminobutyric acid transporter–positive ($VGAT^+$)] and excitatory [vesicular-glutamate transporter 2–positive ($VGLUT2^+$)] presynaptic puncta (Fig. 4A). Significantly more $VGAT^+$ than $VGLUT2^+$ puncta (60% more) were present on the soma of PNN^+ projection neurons (Fig. 4B). Elimination of GAGs from PNNs by injecting AAV-chABC into the lumbar spinal cord did not change the number of excitatory and inhibitory presynaptic terminals on PNN^+ projection neurons (fig. S5, A and B). The number of excitatory and inhibitory presynaptic terminals also remained unchanged 3 days after SNI (fig. S5, C and D).

To study the impact of GAG removal on neuronal activity of lamina I projection neurons, we measured synaptic inputs into these neurons and their intrinsic excitability using whole-cell patch-clamp recording in an ex vivo spinal cord preparation (Fig. 4C) (*26*). Large-diameter (>20 μm) lamina I neurons with mediolateral orientation, which are surrounded by PNNs (fig. S1A), were identified in intact spinal cord for patch clamp recording by using oblique infrared illumination with a light-emitting diode (LED) (*27*). To confirm the presence of PNNs around these neurons, we filled the patched cells with Alexa fluor 488 hydrazide, followed by fixation and aggrecan immunostaining (Fig. 4C). Degradation of GAG with chABC resulted in a depolarization of membrane potential (4.8 ± 1 mV) (Fig. 4, D and E) and an increase in action potential firing rate (Fig. 4, D and F). This effect was specific to neurons with PNNs because chABC did not change membrane potential and firing rate in PNN^- neurons (fig. S5E). Analysis of spontaneous miniature synaptic activity revealed that chABC decreased the frequency of miniature inhibitory postsynaptic currents (mIPSCs) (48% decrease) (Fig. 4, G, H, and I) but not their amplitude (Fig. 4, G and J). No effects on the frequency or amplitude of miniature excitatory postsynaptic currents (mEPSCs) were detected (fig. S6, A to C). Moreover, degradation of PNNs had no effect on passive membrane properties and intrinsic excitability of projection neurons in the presence of synaptic receptor antagonist cocktail (AP-5, DNQX, bicuculline, and strychnine) to suppress both excitatory and inhibitory synaptic activities (fig. S6, D to M). To confirm that disruption of PNNs after nerve injury is accompanied by a decrease in inhibitory inputs, we recorded from lamina I PNN^+ projection neurons on day 3 after SNI and found a significant reduction in the frequency of mIPSCs (Fig. 4, K to M). This reduction was prevented in mice with depletion of microglia (iDTR; $TMEM119^{CreERT2}$) but could be reinstated through removal of PNNs with chABC. Consistent with these results, nerve injury–induced spontaneous pain was alleviated by depletion of microglia (by using the CSF1R inhibitor PLX5622) but was reinstated through removal of PNNs around lamina I projection neurons (Fig. 4N).

Discussion

We have uncovered a mechanism by which activated microglia selectively augment the output of spinal nociceptive circuits and thus evoke pain. A subpopulation of lamina I spinoparabrachial projection neurons are preferentially surrounded by PNNs, which are degraded after peripheral nerve injury in a microglia-dependent manner (fig. S7). Degradation of PNNs is sufficient to enhance projection neuron activity, through reduction of inhibitory synaptic inputs, and to promote pain behaviors.

Whereas in the cortex and hippocampus, PNNs are largely found around parvalbumin interneurons (*28*), in the dorsal horn of the spinal cord, PNNs are found in lamina I exclusively around large-diameter projection neurons and surround various neuronal types in deeper laminae. The selective localization of PNNs around projection neurons, but not other cell types in the superficial dorsal horn, allows for regulation of projection neuron activity, by means of modulation of PNNs, with a high degree of specificity. Because projection neurons are the main output of spinal pain circuits, activation of projection neurons through degradation of PNNs around them is a specific and efficient mechanism to augment the output of spinal nociceptive circuits.

Synaptic terminals are embedded in PNNs, and thus the composition and stability of PNNs have a profound impact on synaptic activity (*5*, *25*). Removal of GAGs that bear negatively charged sulfate groups on PNNs might result in destabilization of synaptic structures and reduce the efficiency of synaptic transmission. Future studies will be required to obtain a better mechanistic understanding of the effect of PNN degradation on inhibitory synapses around projection neurons.

Microglia play crucial roles in the promotion of pain states through several known mechanisms (*1*, *2*). Reduction of inhibitory tone—through microglia-mediated down-regulation of the K^+-Cl^- cotransporter (KCC2), resulting in increased intracellular chloride and attenuated inhibitory inputs—promotes neuropathic pain (*29*). Down-regulation of KCC2 and chloride dysregulation occur in several cell types in lamina I and II at day 7 after nerve injury (*30*, *31*). Microglia-mediated degradation of PNNs selectively augments the activity of a subset of lamina I projection neurons and is observed at 3 days after injury. The existence of several microglia-dependent processes to promote pain that might act in parallel with different cellular specificity and temporal profiles reflects the complexity and robustness of mechanisms underlying neuropathic pain and suggests that efficient treatments should target several processes to reverse hypersensitivity.

Sex differences in the involvement of microglia in mediating chronic pain have been demonstrated (*32*, *33*), although microgliosis occurs equally in both sexes after nerve injury (*32*, *33*), and not all pain models (*34*) and elements of the spinal signaling pathway (*30*) are sexually dimorphic. We found that microglia-dependent degradation of PNNs and its effect on synaptic activity and pain-related behavior is present in both males and females.

Neuronal activity promotes the formation and maintenance of PNNs around parvalbumin-positive neurons (*35*). It remains to be determined how enhanced activity of spinal circuits after peripheral nerve injury affects the dynamics of PNNs around lamina I projection neurons, including their recovery from degradation and long-term maintenance.

Our work uncovers a mechanism by which microglia disinhibit projection neurons after peripheral nerve injury. These findings might lead to the development of therapeutic strategies to reverse neuropathic pain by targeting this newly discovered mechanism.

REFERENCES AND NOTES

1. M. W. Salter, B. Stevens, *Nat. Med.* **23**, 1018–1027 (2017).
2. G. Chen, Y. Q. Zhang, Y. J. Qadri, C. N. Serhan, R. R. Ji, *Neuron* **100**, 1292–1311 (2018).
3. J. W. Fawcett, T. Oohashi, T. Pizzorusso, *Nat. Rev. Neurosci.* **20**, 451–465 (2019).
4. H. H. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9813–9815 (2018).
5. E. Favuzzi *et al.*, *Neuron* **95**, 639–655.e10 (2017).
6. T. Pizzorusso *et al.*, *Science* **298**, 1248–1251 (2002).
7. K. K. Lensjø, M. E. Lepperød, G. Dick, T. Hafting, M. Fyhn, *J. Neurosci.* **37**, 1269–1283 (2017).
8. D. Rowlands *et al.*, *J. Neurosci.* **38**, 10102–10113 (2018).
9. D. Carulli *et al.*, *J. Comp. Neurol.* **494**, 559–577 (2006).

10. R. Wercberger, A. I. Basbaum, *Curr. Opin. Physiol.* **11**, 109–115 (2019).
11. C. Peirs, R. Dallel, A. J. Todd, *J. Neural Transm.* **127**, 505–525 (2020).
12. I. Decosterd, C. J. Woolf, *Pain* **87**, 149–158 (2000).
13. S. D. Shields, W. A. Eckert3rd, A. I. Basbaum, *J. Pain* **4**, 465–470 (2003).
14. T. Buch *et al*., *Nat. Methods* **2**, 419–426 (2005).
15. T. Kaiser, G. Feng, *eNeuro* **6**, ENEURO.0448-18.2019 (2019).
16. S. Jung *et al*., *Mol. Cell. Biol.* **20**, 4106–4114 (2000).
17. S. Lee *et al*., *Am. J. Pathol.* **177**, 2549–2562 (2010).
18. M. K. Zabel *et al*., *Glia* **64**, 1479–1491 (2016).
19. G. Milior *et al*., *Brain Behav. Immun.* **55**, 114–125 (2016).
20. A. A. Staniland *et al*., *J. Neurochem.* **114**, 1143–1157 (2010).
21. D. G. Tervo *et al*., *Neuron* **92**, 372–382 (2016).
22. A. G. J. Skorput *et al*., *PLOS ONE* **17**, e0264938 (2022).
23. J. N. Alves *et al*., *J. Neurosci. Methods* **227**, 107–120 (2014).
24. E. J. Bradbury *et al*., *Nature* **416**, 636–640 (2002).
25. A. Dityatev, M. Schachner, P. Sonderegger, *Nat. Rev. Neurosci.* **11**, 735–746 (2010).
26. P. Szucs, L. L. Luz, D. Lima, B. V. Safronov, *J. Comp. Neurol.* **518**, 2645–2665 (2010).
27. P. Szucs, V. Pinto, B. V. Safronov, *J. Neurosci. Methods* **177**, 369–380 (2009).
28. C. Peirs, R. P. Seal, *Science* **354**, 578–584 (2016).
29. J. A. Coull *et al*., *Nature* **424**, 938–942 (2003).
30. J. C. S. Mapplebeck *et al*., *Cell Rep.* **28**, 590–596.e4 (2019).
31. K. Y. Lee, S. Ratté, S. A. Prescott, *eLife* **8**, e49753 (2019).
32. R. E. Sorge *et al*., *Nat. Neurosci.* **18**, 1081–1083 (2015).
33. S. Taves *et al*., *Brain Behav. Immun.* **55**, 70–81 (2016).
34. J. Peng *et al*., *Nat. Commun.* **7**, 12029 (2016).
35. G. Devienne *et al*., *J. Neurosci.* **41**, 5779–5790 (2021).

ACKNOWLEDGMENTS

Funding: This work was supported by Canadian Institutes of Health Research PJT-162412 (A.Kh.); Rita Allen Foundation Award in Pain (A.Kh.); FRN (154281) (J.S.M.); and the Canada Research Chair in Chronic Pain and Related Brain Disorders (Y.D.K.). **Author contributions:** Conceptualization: S.T., J.S.M., A.Kh.; Methodology: S.T., N.G., A.U.G., E.M.-P., C.W., N.Y., W.C., J.H., R.B.R., N.W., A.C., G.L., E.M.M., C.G.G.; Investigation: S.T., N.G., A.U.G., E.M.-P., C.W., N.Y., W.C., K.C.L.; Funding acquisition: J.S.M. and A.Kh.; Supervision: A.Kh., M.P.-K., J.Z., J.W.F., L.D., A.R.-d.-S., Y.D.K., J.S.M., A.Kh.; Writing: S.T., J.S.M., A.Kh. **Competing interests:** None declared. J.H. is also affiliated with the University of British Columbia, Faculty of Medicine. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl6773
Materials and Methods
Figs. S1 to S7
References (*36–45*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 28 July 2021; accepted 16 May 2022
Published online 26 May 2022
10.1126/science.abl6773

REVIEW SUMMARY

GUT PHYSIOLOGY

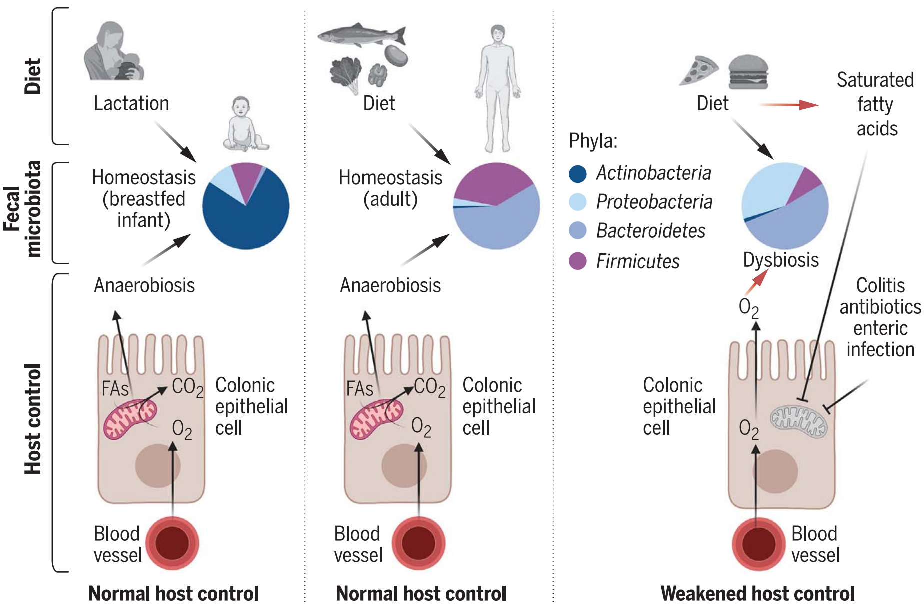
The microbiome and gut homeostasis

Jee-Yon Lee, Renée M. Tsolis, Andreas J. Bäumler*

BACKGROUND: Changes in the composition of gut-associated microbial communities (gut microbiota) are linked to a wide spectrum of human diseases, such as cancer, obesity, and even neurological disorders. Understanding the factors that shape the gut microbiota is therefore a primary objective of microbiome research. One factor governing the composition of the gut microbiota is diet. Weaning marks an abrupt shift in diet, which is associated with phylum-level changes in the fecal microbiota composition. But the fecal microbiota of healthy adults and that of breastfed infants before weaning are both homeostatic communities. Thus, diet-induced changes in the composition of the gut microbiota are not always associated with disease. Furthermore, large variation in the species composition of the gut microbiota between individuals makes it impossible to define homeostasis or dysbiosis by the presence or absence of specific microorganisms. Limited information on the ecological causes of dysbiosis and the causative effects of dysbiosis on disease makes it difficult to translate research into medical interventions.

ADVANCES: In this Review, we explore whether homeostasis and dysbiosis can be better defined by looking at a second factor that governs gut microbiota composition and function: the host environment. Bacterial growth requires energy in the form of adenosine triphosphate (ATP). Microbes that can maximize energy production in a specific environment will accelerate their growth and dominate microbial communities. Because respiration yields more energy than fermentation, bacteria that respire will dominate a microbial community when respiratory electron acceptors, such as oxygen, are present in the environment. The host uses these principles of bacterial community composition to shape its gut microbiota. Our body maintains a high luminal oxygen concentration in the small intestine to drive the microbial community composition to be dominated by bacteria that use respiration for energy production. By contrast, through maintaining the epithelium in a state of physiological hypoxia, the host limits access to oxygen in the lumen of the large intestine, which drives a dominance of bacteria that use fermentation for energy production. Recent insights into the ecological drivers of dysbiosis suggest that a disruption of homeostasis is generally associated with a weakening of host control mechanisms that regulate the microbiota. Host control mechanisms that limit the availability of oxygen for bacteria in the colon become weakened in mouse models of colorectal cancer, ulcerative colitis, antibiotic treatment, or enteric infection. The resulting increase in the availability of oxygen increases the abundance of bacteria that respire oxygen in the colonic microbiota. Similarly, negative health effects of a Western-style high-energy, low-fiber diet are associated with a weakening of host factors that control the microbial environment in the gut. These mechanistic insights suggest that dysbiosis is associated with a state of weakened host control over the microbial environment. Conversely, gut homeostasis represents a state where host functions that control microbial growth are normal (i.e., characteristic of a healthy or normally functioning individual).

OUTLOOK: The idea that dysbiosis is characterized by an underlying impairment of host functions that regulate the gut microbiota and microbial metabolism suggests that measurements of these host functions could provide a more quantitative insight into the state of microbiome homeostasis than current microbe-centric approaches. For example, measurements of oxygen concentrations along the longitudinal axis of the intestine could be used to determine a normal range that is characteristic of a healthy or normally functioning individual. Measurements within the normal range might indicate homeostasis, whereas values outside this range would indicate dysbiosis. Furthermore, targeting the host functions that regulate the gut microbiota with therapeutic approaches could fuel the conception of new strategies to remediate dysbiosis with drugs that restore host control over the microbiota. ■



Dysbiosis is linked to weakened host control over microbial growth. The composition of the gut microbiota is governed by diet and by host control over microbial growth, which includes high mitochondrial oxygen consumption in the colonic epithelium to maintain anaerobiosis. The transition between nursing (left) and adult-type diet (middle) shifts the microbiota composition, but both states represent homeostasis. Dysbiosis ensues only when host control of microbial growth is weakened because dietary factors, antibiotics, or underlying host conditions reduce oxygen consumption in the colonic epithelium, which drives dysbiosis by increasing the diffusion of oxygen into the intestinal lumen (right). FAs, fatty acids; O_2, oxygen; CO_2, carbon dioxide [created using BioRender.com].

Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, Davis, CA 95616, USA.
*Corresponding author. Email: ajbaumler@ucdavis.edu
Cite this article as J.-Y. Lee *et al.*, *Science* 377, eabp9960 (2022). DOI: 10.1126/science.abp9960

READ THE FULL ARTICLE AT https://doi.org/10.1126/science.abp9960

REVIEW

GUT PHYSIOLOGY

The microbiome and gut homeostasis

Jee-Yon Lee, Renée M. Tsolis, Andreas J. Bäumler*

Changes in the composition of the gut microbiota are associated with many human diseases. So far, however, we have failed to define homeostasis or dysbiosis by the presence or absence of specific microbial species. The composition and function of the adult gut microbiota is governed by diet and host factors that regulate and direct microbial growth. The host delivers oxygen and nitrate to the lumen of the small intestine, which selects for bacteria that use respiration for energy production. In the colon, by contrast, the host limits the availability of oxygen and nitrate, which results in a bacterial community that specializes in fermentation for growth. Although diet influences microbiota composition, a poor diet weakens host control mechanisms that regulate the microbiota. Hence, quantifying host parameters that control microbial growth could help define homeostasis or dysbiosis and could offer alternative strategies to remediate dysbiosis.

Host-associated microbial communities, the microbiota, play a central role in human health (*1*). Many human illnesses are linked to an imbalance in microbial communities, termed dysbiosis, which is commonly defined as a decrease in microbial diversity, the presence of potentially harmful microbes, or the absence of beneficial ones (*2*). Understanding factors that influence microbiota composition and function is therefore a fundamental goal of microbiome research (*3*). To approach this problem, most investigators probe the microbiota to ask what is there (e.g., using microbiota profiling), what could they do (e.g., using metagenomics), or what are they doing (e.g., using metatranscriptomics, proteomics, or metabolomics) (*4*). However, microbial communities show high diversity between individuals (*5*), which makes it challenging to define what constitutes a balanced microbial community in a healthy state by simply cataloging bacterial taxa (*6*–*9*). As a result, homeostasis cannot be explained by the presence or absence of specific microbial species, which also challenges the validity of defining dysbiosis using this approach (*7*). This gulf in understanding and the complexity of microbial communities make it impossible to translate analysis of the microbiota into a unit of measurement that could be used to quantify homeostasis or dysbiosis (*10*).

In this Review, we examine whether the concept that the host controls the environment to ensure a microbiota composition that sustains health (*11*–*17*) could provide a way of moving past the current impasse on how to quantify homeostasis. According to this view, host functions controlling environmental parameters that are important for microbial growth on body surfaces could be viewed as part of an immune system subdivision that is tasked with maintaining microbiome homeostasis (*15*, *18*, *19*). We discuss whether such host functions could be measured to quantify homeostasis and dysbiosis. Our analysis is focused on the intestinal microbiota because imbalance in this microbial community is linked to many human maladies in the gut and systemically (*1*). However, to establish the ecological ground rules, we will start out by reviewing some fundamental mechanisms that shape microbial communities in other environments.

Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, Davis, CA 95616, USA.
*Corresponding author. Email: ajbaumler@ucdavis.edu

Redox reactions select successful metabolic strategies in microbial communities

Bacterial growth requires energy in the form of adenosine triphosphate (ATP) to power biosynthetic processes (*20*). ATP is generated through redox reactions that involve the transfer of electrons from an electron donor, such as glucose, to an electron acceptor, such as oxygen (O_2). On the basis of the redox reactions used for ATP production, bacteria can be classified into metabolic groups, each containing microbes that exploit the same class of resources in a similar way (Fig. 1) (*21*). Examples of metabolic groups include the facultatively anaerobic bacteria, which produce ATP using oxygen or nitrate (NO_3^-) but are capable of switching to fermentation when these electron acceptors are absent. A second metabolic group important for this Review is the obligately anaerobic primary fermenters, which cannot grow under

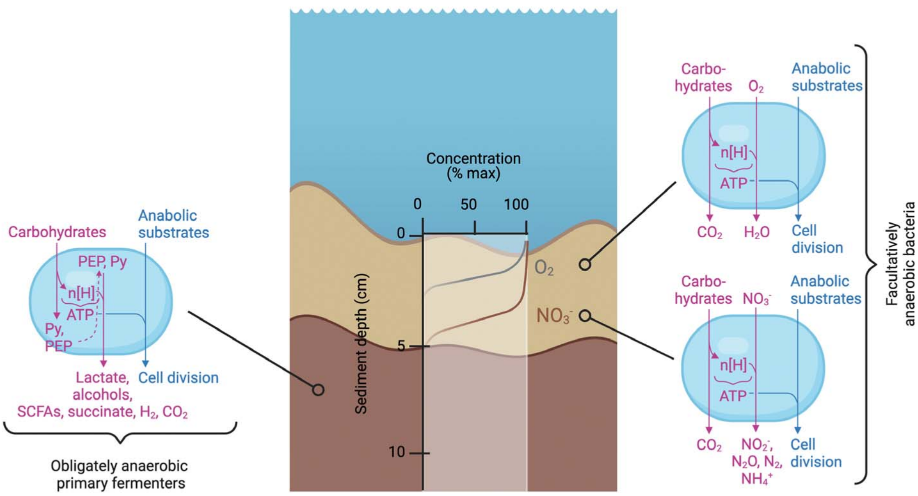


Fig. 1. Redox reactions that control bacterial community composition in lake sediments. Bacteria generate energy (ATP) through redox reactions (magenta font and arrows) that involve the transfer of electrons (n[H]) from an electron donor (e.g., a carbohydrate) to either an exogenous electron acceptor, such as oxygen (O_2) or nitrate (NO_3^-), or to an endogenous electron acceptor, such as phosphoenolpyruvate (PEP) or pyruvate (Py). ATP production, which is coupled to these redox reactions through substrate-level phosphorylation or oxidative phosphorylation, powers synthesis of polymers (blue font and arrows) required for bacterial growth (cell division). Bacteria using the redox reaction with the greatest free energy prevail, which determines which metabolic groups of bacteria can dominate microbial communities in a lake sediment (center). Facultatively anaerobic bacteria that use oxygen, the electron acceptor that generates the greatest free energy, dominate the upper layer of a lake sediment (right, top). Once oxygen has been depleted, facultative anaerobic bacteria respire nitrate to maintain their dominance (right, bottom). Oxygen and nitrate are depleted in deeper sediment layers, which renders obligately anaerobic primary fermenters that use endogenous electron acceptors the most successful metabolic strategy (left). CO_2, carbon dioxide; H_2O, water; NO_2^-, nitrite; N_2O, nitrous oxide; N_2, nitrogen; NH_4^+, ammonium; SCFAs, short-chain fatty acids; H_2, hydrogen. Panels on the right and left were modified from a presentation by A. Spormann with permission [created with BioRender.com].

atmospheric oxygen concentrations and generate ATP by oxidizing available carbohydrates using endogenous electron acceptors, such as pyruvate or phosphoenolpyruvate. Microbes that can maximize ATP production in a specific environment will accelerate their growth and dominate microbial communities. As a result, redox reactions that generate the greatest free energy are the principal drivers of bacterial community composition in nature (*22*, *23*).

For example, oxygen, the electron acceptor with the highest redox potential, is used by facultatively anaerobic bacteria near the surface of lake sediments (*24*) (Fig. 1). Once oxygen has been depleted, facultatively anaerobic bacteria resort to oxidizing organic compounds using nitrate, an electron acceptor that is second only to oxygen in its redox potential (*25*). The predominant usage of oxygen or nitrate as electron acceptors in the upper layers of lake sediment drives the microbial community composition to be dominated by facultatively anaerobic bacteria (*24*, *25*). Conversely, depletion of oxygen and nitrate in deeper sediment layers renders redox reactions that use endogenous electron acceptors, such as pyruvate or phosphoenolpyruvate, the predominant metabolic strategy. Consequently, obligately anaerobic primary fermenters dominate microbial communities in the lower layers of lake sediments (*26*). In brief, the depth-related changes in the composition of microbial communities in lake sediments are governed by the availability of respiratory electron acceptors (*24*, *25*).

The concept that redox reactions are an important driver of bacterial community composition also applies to the gut microbiota (*13*). Similar to the depth-related changes in oxygen availability in lake sediments, oxygen levels in the gut decrease from the duodenum to the large intestine (*27*) (Fig. 2). As a result, the low-density microbial community in the small intestine is dominated by facultatively anaerobic bacteria belonging to the class *Bacilli* (phylum *Firmicutes*) and the order *Enterobacterales* [ord. nov. (*28*), phylum *Proteobacteria*], whereas obligately anaerobic primary fermenters belonging to the classes *Bacteroidia* (phylum *Bacteroidetes*) and *Clostridia* (phylum *Firmicutes*) dominate the high-density microbial community in the large intestine (*29*, *30*). The colonic microbiota is of particular interest for human health (*1*) because this absorptive organ harbors ~100-fold more microbes than any other microbial community in our body (*31*), which makes it the principal source of microbial metabolites. In turn, metabolites produced by the colonic microbiota profoundly influence host physiology, which makes changes in the microbial metabolite profile a common contributor to human disease (*32*–*37*). A high-density microbial community in the colon is beneficial because food that escapes

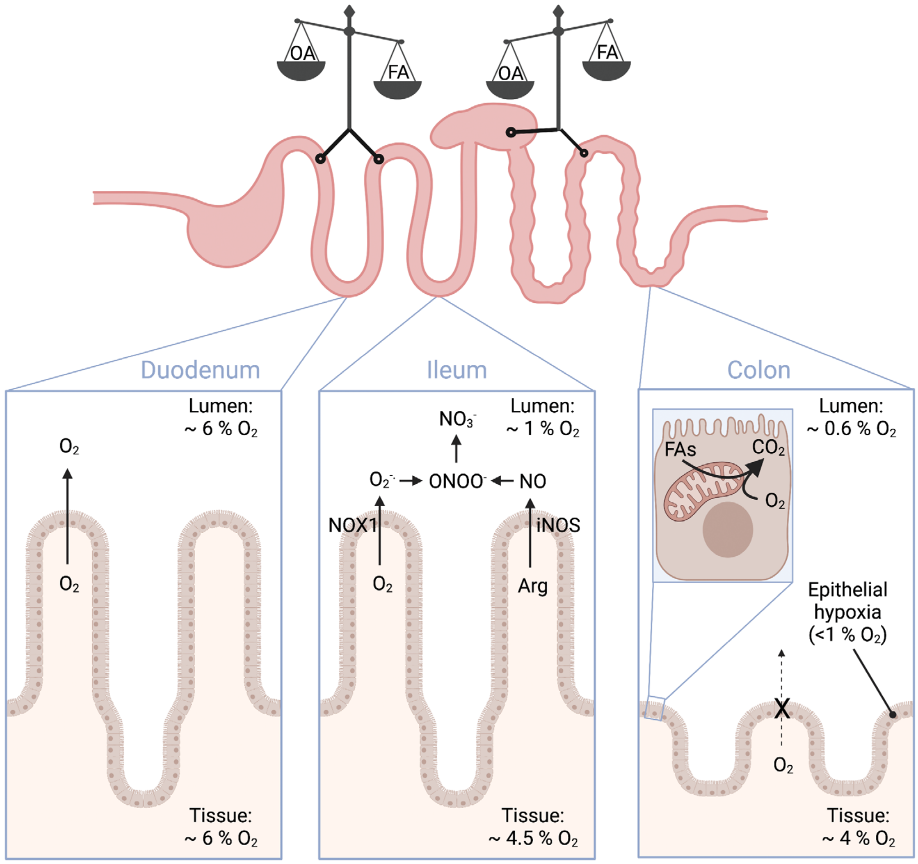


Fig. 2. The host controls electron acceptor availability in the intestine. The murine intestine (middle schematic) features a longitudinal gradient of luminal oxygen levels declining from the duodenum to the colon. In the duodenum (bottom left), luminal and intravascular oxygen levels are similar, which suggests a free exchange of oxygen between tissue and lumen. By contrast, in the colon (bottom right), the epithelium is maintained in a state of physiological hypoxia (<1% O_2) because oxygen is consumed in the mitochondria through oxidative phosphorylation, which is coupled to oxidation of fatty acids (FAs). In turn, epithelial hypoxia limits diffusion of oxygen into the intestinal lumen. As a result, the intravascular oxygen levels in the colon are much higher than the luminal oxygen levels. In addition to controlling luminal oxygen levels, the host also controls the availability of nitrate in the intestinal lumen. The ileum (bottom middle) features epithelial synthesis of an epithelial NADPH oxidase (NOX1) and an inducible nitric oxide synthase (iNOS). Superoxide ($O_2^{\cdot -}$) and nitric oxide (NO) generated by these enzymes react to form peroxynitrite ($ONOO^-$), which decomposes to nitrate in the ileal lumen. The presence of oxygen and nitrate in the small intestine drives a dominance of facultatively anaerobic bacteria (FA) over obligately anaerobic bacteria (OA) (top schematic). Conversely, the absence of oxygen and nitrate in the large intestine favors OA over FA [created with BioRender.com].

digestion by host enzymes in the small intestine is converted by the colonic microbiota into fermentation products that are absorbed by the host for nutrition (*38*, *39*). Conversely, mechanisms that maintain a low microbial density in the small intestine are beneficial for the host because small intestinal bacterial overgrowth is associated with excessive deconjugation of bile, which can cause malabsorption, diarrhea, and weight loss (*40*).

The host selects which metabolic groups dominate the gut microbiota

Analogously to lake sediments, it is commonly assumed that microbial depletion of oxygen is also the main factor regulating electron acceptor availability in the intestine. Accordingly, conventional wisdom holds that facultatively anaerobic bacteria generate an oxygen gradient by gradually depleting this electron acceptor as contents pass through the small intestine, thereby preserving anaerobiosis in the large intestine (*41*, *42*). Furthermore, because facultatively anaerobic *Bacilli* and *Enterobacterales*, such as *Escherichia coli*, are commonly among the first species to colonize the infant gut (*43*–*45*), it is a widely held belief that oxygen consumption by these taxa is required to pave the way for the subsequent engraftment of obligately anaerobic *Bacteroidia* and *Clostridia*

during microbiota assembly (41, 42). However, the following observations challenge the dogma that depletion of oxygen by facultatively anaerobic bacteria is the main factor controlling electron acceptor availability in the intestine.

First, comparison of luminal oxygen levels in the duodenum, ileum, and cecum shows that similar oxygen partial pressures prevail in comparable segments of the gut of conventional mice compared with germ-free mice. This work indicates that the notable decrease in luminal oxygen availability in the cecum compared with the duodenum is maintained independently of the microbiota (27).

Second, engraftment of germ-free mice with *E. coli* facilitates subsequent colonization with *Bacteroides thetaiotaomicron*, but this process does not require *E. coli* to deplete oxygen (46). Although *E. coli* lowers the implantation dose for *B. thetaiotaomicron*, a similar decrease in the *B. thetaiotaomicron* implantation dose is observed after engrafting germ-free mice with an *E. coli hemA* mutant, which cannot respire oxygen (46). These data suggest that colonization of the neonate gut with facultatively anaerobic *Proteobacteria* promotes a subsequent engraftment of obligately anaerobic *Bacteroidia* through mechanisms that are independent of microbial oxygen depletion.

Third, luminal oxygen levels in the duodenum of germ-free mice are close to tissue levels, which suggests that oxygen is derived from the circulation and freely diffuses from the tissue into the duodenal lumen (27) (Fig. 2). By contrast, the host limits diffusion of oxygen into the large bowel by maintaining a state of physiological hypoxia (<1% O_2) in the large intestinal epithelium (14, 47). Through this mechanism, the host regulates luminal oxygen concentration in the cecum to ~0.6%—effectively an anaerobic state that is also preserved in the cecum of germ-free mice (27). Together, these data indicate that the oxygen gradient along the longitudinal axis of the intestinal lumen is generated and maintained by the host, not by microbial oxygen depletion.

The luminal oxygen level of ~6% (45 mmHg O_2) found in the duodenum of mice is considerably higher than that in the ileum, which averages only 1% oxygen (7.5 mmHg) (27). Although oxygen availability is limited, epithelial cells of the ileum constitutively synthesize a reduced form of nicotinamide adenine dinucleotide phosphate (NADPH)–oxidase, termed NOX1 (48), and an inducible nitric oxide synthase, termed iNOS (49). Superoxide ($O_2^{\bullet -}$) produced by NOX1 reacts with nitric oxide (NO·) formed by iNOS to generate peroxynitrite ($ONOO^-$) (50, 51), which decomposes to nitrate in the intestinal lumen (52) (Fig. 2). As a result, the nitrate concentration in the ileal lumen of mice averages ~6 mM during gut homeostasis (53). By contrast, host-derived nitrate is not available in the lumen of the duodenum or cecum (53) because iNOS is not synthesized in these tissues (49). Notably, mice deficient for the synthesis of iNOS and NOX1 exhibit an increase in the abundance of obligately anaerobic bacteria in the ileal microbiota, such that the composition of this microbial community closely resembles that in the colon (48). These observations are consistent with the idea that host-derived nitrate helps to maintain a dominance of facultatively anaerobic bacteria in the distal small intestine.

To conclude, the host maintains a high luminal oxygen concentration in the proximal small intestine (27) and delivers nitrate to the lumen of the distal small intestine (53). The predominant usage of oxygen or nitrate as electron acceptors in the small intestine drives the microbial community composition to be dominated by facultatively anaerobic bacteria (30). By contrast, through limiting access to oxygen and nitrate in the large intestine (54), the host drives a dominance of obligately anaerobic primary fermenters, which commonly include members of the classes *Bacteroidia* and *Clostridia* (55).

The picture emerging from these studies is that the bacterial community composition in the intestine is shaped by the same thermodynamically predictable hierarchy of redox reactions that also govern the depth-related changes in the composition of microbial communities in lake sediments. However, unlike in lake sediments, where microbial depletion of electron acceptors controls their availability (24, 25), the availability of oxygen and nitrate in the intestine is firmly under host control through the regulation of epithelial hypoxia and iNOS synthesis (54). By differentially controlling the availability of oxygen and nitrate along the longitudinal axis of the intestine, the host controls which microbial metabolic strategies prevail in the small versus the large intestine by using the same thermodynamic principles that govern bacterial community composition in nature.

Diet shapes the gut microbiota composition during homeostasis

When the gut is in homeostasis, the host maintains concentrations of oxygen and nitrate at constant levels in each part of the intestine (27, 53). However, changes in the availability of diet-derived electron donors can cause variation in the composition of the microbiota. In the gastrointestinal tract, foodstuffs are broken down by host enzymes in the upper gastrointestinal tract into simple sugars, amino acids, fatty acids, and diglycerides. These nutrients are efficiently absorbed in the small intestine and do not represent a major source of electron donors for the colonic microbiota. However, indigestible polymers, such as plant polysaccharides (fiber) or cartilage-derived glycans, are available as diet-derived electron donors for the colonic microbiota (56). A second class of diet-derived electron donors in the colon are simple sugars that are poorly absorbed in the small intestine, which fall under the umbrella of FODMAPs, an acronym for fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (57). Because the host limits the availability of the exogenous electron acceptors oxygen and nitrate in the colonic lumen during homeostasis, FODMAPs and indigestible polymers are broken down by obligately anaerobic primary fermenters using endogenous electron acceptors (Fig. 1). In turn, ATP generated through these redox reactions drives a dominance of obligately anaerobic primary fermenters in the colonic microbiota during homeostasis.

Cultural influences cause considerable variation in dietary behavior between individuals. As a result, individual variation in feeding behavior results in variation in diet-derived electron donors, which translates into personal variation in human fecal microbiota composition (58–60). The most abrupt and pronounced change in dietary behavior is associated with weaning. In mammals, nutrition of the infant is under maternal control through lactation. Milk contains fat and proteins that are digested by host enzymes and absorbed in the small intestine of the infant. Additionally, milk contains FODMAP-like oligosaccharides, which are poorly absorbed in the small intestine, do not directly provide nutrition for the infant, and serve as the main source of electron donors for the colonic microbiota (61, 62). Milk oligosaccharides are a heterogeneous mix of soluble glycans with oligosaccharide compositions that differ between mammalian species (63). As the infant microbiota matures, niches carved out by human milk oligosaccharides become occupied by *Bifidobacterium longum* subspecies *infantis* (phylum *Actinobacteria*), a microbe that efficiently degrades these electron donors (64, 65). Consequently, this obligately anaerobic primary fermenter accounts for up to 70% of the fecal 16*S* ribosomal RNA (rRNA) amplicon sequencing reads in breastfed infants (66) (Fig. 3). Weaning removes human milk oligosaccharides from the diet, which leads to an extinction of *B. longum* subspecies *infantis* from the fecal microbiota (67). After weaning, the electron donor menu for the colonic microbiota changes abruptly, as human milk oligosaccharides are replaced by diet-derived FODMAPs and indigestible polymers. A large repertoire of genes encoding polysaccharide lyases and glycosyl hydrolases make *Bacteroidia* well suited to degrade plant polysaccharides (68), whereas *Clostridia* are the main consumers of diet-derived FODMAPs during homeostasis (69). As a result, diet-derived electron donors drive the colonic microbiota composition after weaning toward a dominance of obligately anaerobic primary

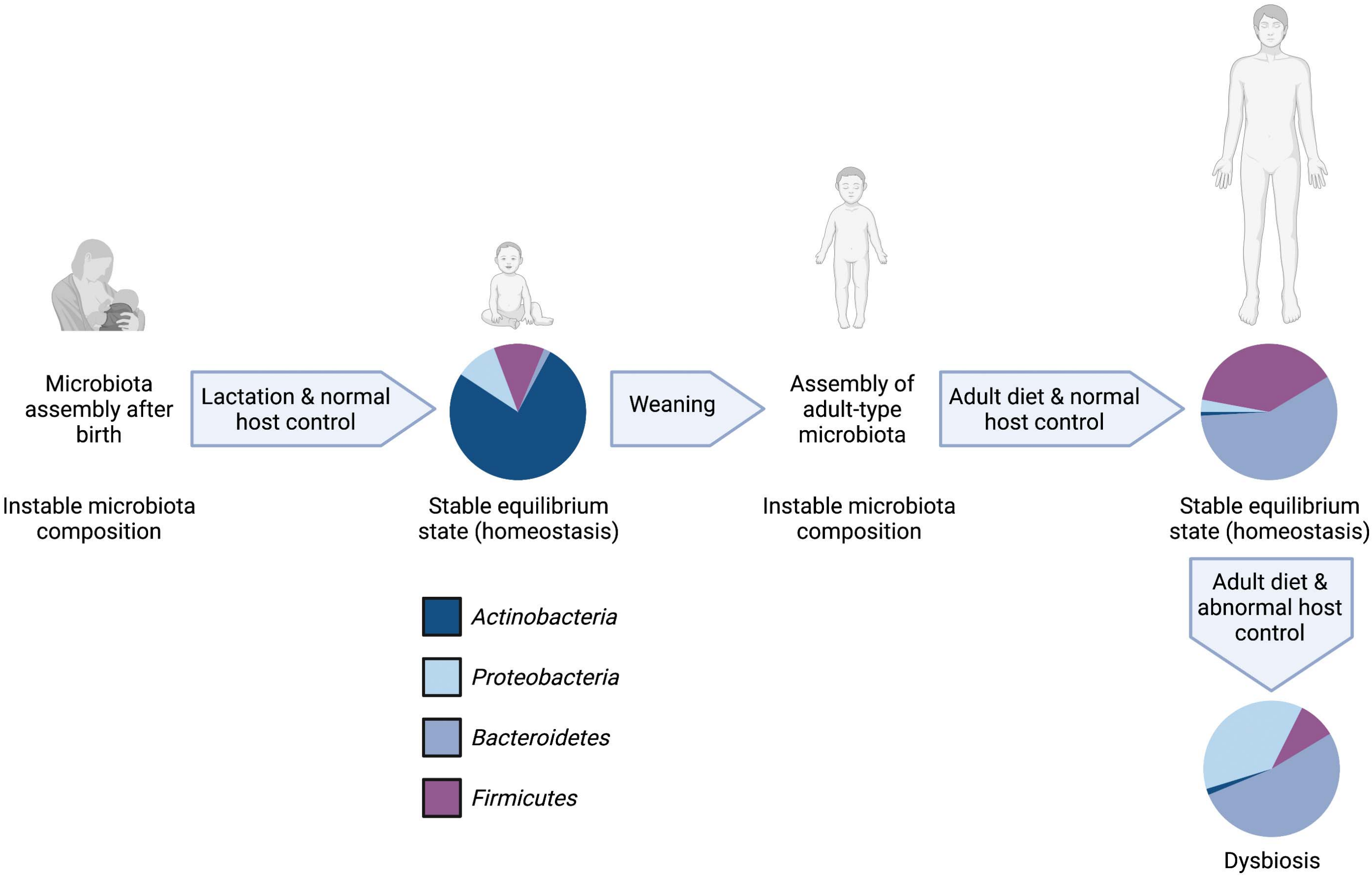


Fig. 3. Fecal microbiota assembly and homeostasis. After birth, fecal microbiota assembly is associated with fluctuations in the species composition. Once ecological niches created by lactation and host-derived resources are occupied, the fecal microbiota enters an equilibrium state characterized by a dominance of *Actinobacteria*. Weaning opens ecological niches by introducing new diet-derived electron donors, which initiates a second round of microbiota assembly associated with a brief (i.e., 2- to 3-week) instability of the microbiota composition. Diet and host factors that control microbial growth conditions act in concert to drive a stable equilibrium state, in which the adult-type microbiota is dominated by *Firmicutes* and *Bacteroidetes*. An infant-type microbiota dominated by *Actinobacteria* and an adult-type microbiota dominated by *Firmicutes* and *Bacteroidetes* both represent homeostatic microbial communities because host control over the microbial environment is normal. A weakening of host functions that control parameters important for microbial growth generates abnormal growth conditions that lead to dysbiosis. These abnormal growth conditions often involve an increase in the availability of oxygen and nitrate in the colonic lumen, which can be caused by high-fat diet, antibiotic usage, intestinal inflammation, enteric infection, or colorectal cancer. Abnormally high levels of these host-derived electron acceptors in the colon drive dysbiosis in the fecal microbiota that is characterized by an elevated abundance of *Proteobacteria* [created with BioRender.com].

fermenters belonging to the classes *Bacteroidia* and *Clostridia* (29, 30).

The ecological succession during weaning, from a colonic microbiota dominated by *Actinobacteria* to an adult-type microbiota dominated by *Bacteroidia* and *Clostridia* illustrates that diet-derived electron donors have a marked influence on the gut microbiota composition (67). After weaning, new ecological niches carved out by diet-derived FODMAPs and indigestible polymers become successively occupied by *Bacteroidia* and *Clostridia* species. While this assembly process is in progress, susceptibility to enteric bacterial pathogens increases transiently (70). However, once the assembly process is complete, an adult-type microbiota can prevent engraftment of new microbes through priority effects, thereby providing resistance to infection with enteric pathogens (71). Notably, both an *Actinobacteria*-dominated microbiota in breastfed infants and a microbiota dominated by *Bacteroidia* and *Clostridia* in adults represent homeostatic microbial communities (Fig. 3).

Causes and consequences of dysbiosis

Whereas the repertoire of diet-derived electron donors determines which redox reactions are available for microbial growth during homeostasis, an increased intake of a high-energy (i.e., high-fat, high-sucrose and/or -fructose), low-fiber diet (i.e., a diet characterized by a lack of whole grains and vegetables) can contribute to a disruption of homeostasis (72). A high-fat diet contains foods with high fat content (e.g., processed meats or animal fat), resulting in more than 35% of caloric intake being derived from fat. Examples of foods high in sucrose or fructose include candy, desserts, sugary soft drinks, fruit juices, and commercial cereals. A sedentary lifestyle in conjunction with increased high-energy, low-fiber food consumption has generated a positive energy imbalance in high-income countries, which fuels obesity rates (73, 74). Obesity is associated with dysbiosis, which is in turn associated with accelerated weight gain (75) and exacerbated obesity-associated states, such as cardiovascular disease (76), colitis (77), and colorectal cancer (78). Notably, high-energy, low-fiber food disrupts homeostasis by directly or indirectly impairing host control mechanisms that shape the environment of the gut microbiota.

An aspect of high-energy food that indirectly disrupts host control over the microbial environment is a lack of indigestible polymers in

the diet. A decreased diversity in dietary plant polysaccharides, as shown in a study of immigration to the US from Thailand, is linked to a displacement of *Prevotella* species (phylum *Bacteroidetes*) by *Bacteroides* species (phylum *Bacteroidetes*) in the fecal microbiota (*59*). Whereas *Prevotella* species specialize in catabolizing diet-derived fiber, *Bacteroides* species are fiber eaters that switch to breaking down host-derived mucin glycans when plant polysaccharides are removed from the diet of laboratory mice (*79*, *80*). The resulting degradation of mucin glycans can then lead to erosion of the colonic mucus barrier (*81*). The host maintains an inner colonic mucus layer containing a low density of microbes, whereas the outer mucus layer is densely colonized by microorganisms (*82*). This organization into inner and outer mucus layers generates cross-sectional heterogeneity in the density of colonic microbial communities. The low bacterial density in the inner mucus layer is thought to protect the epithelial surface from the high-density microbial communities inhabiting the outer mucus layer and the lumen. Consistent with this idea, erosion of the inner mucus layer increases susceptibility to infection in a mouse model (*81*) and is associated with penetration of bacteria into crypts during chemically induced colitis in mice (*83*) and in ulcerative colitis patients (*84*, *85*). In short, a low-fiber diet drives erosion of the mucus barrier, which impairs host control of the colonic microbiota.

A second aspect of high-energy, low-fiber food is its high sucrose and/or high fructose content. Overconsumption of these simple sugars can overwhelm absorption in the small intestine, resulting in transit into the colon (*86*), where these electron donors trigger changes in the microbiota composition by increasing the abundance of *Firmicutes* that consume sucrose or fructose (*87*, *88*). A long-term high-fructose diet promotes the growth of mucin glycan–degrading bacterial species (*89*). It also impairs the colonic mucus barrier (*89*) and increases susceptibility to colitis (*90*). However, the chain of events through which a high-sugar diet weakens the mucus barrier remains incompletely understood.

A third aspect of high-energy, low-fiber food is its high fat content, which directly impairs host cell functions that shape the microbial environment. A prolonged high-fat diet induces low-grade colonic inflammation (*91*), which can be exacerbated by antibiotic treatment (*92*) or by consumption of large amounts of sucrose (*93*). Mechanistically, saturated fatty acids, a component of a high-fat diet, trigger oxidative stress (*91*) by inducing elevated mitochondrial hydrogen peroxide production (*94*, *95*). As a result, saturated fatty acids impair mitochondrial bioenergetics in the intestinal epithelium if a high-fat diet is prolonged (*96*). Reduced mitochondrial bioenergetics trigger endoplasmic reticulum stress in mucus-producing goblet cells, thereby weakening the mucus barrier by reducing mucus production (*91*). Because mitochondrial oxygen consumption maintains the colonic surface in a state of physiological hypoxia (*54*), another result of mitochondrial impairment by a high-fat diet is an increase in epithelial oxygenation, which in turn raises oxygen availability in the colonic lumen (*92*, *96*). Furthermore, reduced mitochondrial activity in colonic epithelial cells induces iNOS synthesis (*54*). Consequently, a diet rich in saturated fatty acids leads to epithelial release of nitric oxide in the colon (*96*). Superoxide produced by NOX1, which is synthesized constitutively in the colonic epithelium (*48*), reacts with nitric oxide to form nitrate in the intestinal lumen (*96*). In summary, a high-fat diet rich in saturated fatty acids directly acts on the colonic epithelium to impair the host's ability to limit the availability of oxygen and nitrate to microbes in the colonic lumen.

Host control over electron acceptor availability is important for balancing the microbiota composition to maintain homeostasis (*12*, *13*, *19*). Conversely, impairment of these host control mechanisms can drive dysbiosis (*14*, *17*, *97*). Conditions that increase the availability of host-derived oxygen and nitrate in the colonic lumen cause dysbiosis by fueling growth of facultatively anaerobic bacteria in the colonic microbiota (*97*–*99*). Levels of host-derived oxygen and nitrate become elevated in the colonic lumen not only by a high-fat diet (*96*), but also after antibiotic treatment (*54*, *100*); after infection with enteric pathogens (*101*–*108*); and in mouse models of ulcerative colitis (*109*–*112*), cancer cachexia (*113*), or colorectal cancer (*114*, *115*). These diseases are all marked by an increased abundance of facultatively anaerobic bacteria in the colonic microbiota because such organisms can use oxygen or nitrate as respiratory electron acceptors to fuel their growth (*54*, *92*, *96*, *100*–*115*) (Fig. 1). An elevated abundances of facultatively anaerobic bacteria in the colonic microbiota is a signature of microbial dysbiosis (*116*). However, this change in the microbiota composition is secondary to an underlying impairment of host control mechanisms that limit the availability of oxygen and nitrate in the colonic lumen (*17*). Mechanistically, host control over the availability of oxygen and nitrate in the colon can become impaired by compounds that act directly on epithelial cells, as is the case for saturated fatty acids (*96*), or by compounds that act directly on microbes, such as antibiotics. Antibiotic-mediated disruption of the colonic microbiota depletes short-chain fatty acids, which in turn triggers a shift in epithelial metabolism to increase the availability of oxygen and nitrate in the colonic lumen (*54*). These mechanistic insights suggest that dysbiosis is associated with a state of weakened host control over the microbial environment. Conversely, gut homeostasis represents a state where host functions that control microbial growth are normal (i.e., characteristic of a healthy or normally functioning individual) (*71*) (Fig. 3).

There is mounting evidence that an increased availability of host-derived oxygen and nitrate in the colon does not merely change which bacteria are there, but also what facultatively anaerobic bacteria are doing. Bacteria have proteinaceous structures in their cytoplasm that are analogous to the membrane-bound organelles of eukaryotic cells. These microcompartments typically encase enzymes involved in catabolic pathways that generate a toxic metabolic intermediate, such as propionaldehyde or acetaldehyde (*117*). Facultatively anaerobic bacteria produce these aldehyde intermediates during the catabolism of 1,2 propanediol, ethanolamine, or choline in microcompartments (*117*). Notably, facultatively anaerobic bacteria require a respiratory electron acceptor to power this catabolism (*96*, *118*–*122*) (Fig. 4A). Therefore, microcompartments do not provide a growth advantage in the colon during gut homeostasis when oxygen and nitrate are scarce. However, conditions that increase the availability of exogenous electron acceptors in the colonic lumen enable facultatively anaerobic *Enterobacterales* to initiate catabolism of ethanolamine, 1,2 propanediol, and choline (*96*, *119*, *120*) (Fig. 4, B and C). This metabolic shift is relevant because microcompartment-derived metabolites can adversely affect the health of the host (*123*).

Trimethylamine (TMA) produced during the catabolism of choline—a component of high-energy, low-fiber food—is perhaps the microcompartment-derived metabolite that is most relevant for human health because it has been linked to cardiovascular disease (*76*) (Fig. 4C). Gene clusters responsible for TMA production are commonly found in obligately anaerobic *Clostridia* and facultatively anaerobic *Enterobacterales* (*124*, *125*), but only the latter taxon features an elevated relative abundance in the fecal microbiota of individuals on a high-fat diet (*126*–*128*). In addition to increasing the abundance of *Enterobacterales*, saturated fatty acids heighten the availability of oxygen and nitrate in the colon. In turn, these electron acceptors power microcompartment-dependent choline catabolism, which turns *Enterobacterales* into dominant producers of TMA in the microbiota (*96*) (Fig. 5). TMA is absorbed by the host and converted by flavin monooxygenases in the liver to trimethylamine *N*-oxide (TMAO), which can be measured in the circulation (*129*, *130*). Elevated TMAO concentrations in the blood are linked to cardiovascular disease (*76*), increasing the relative

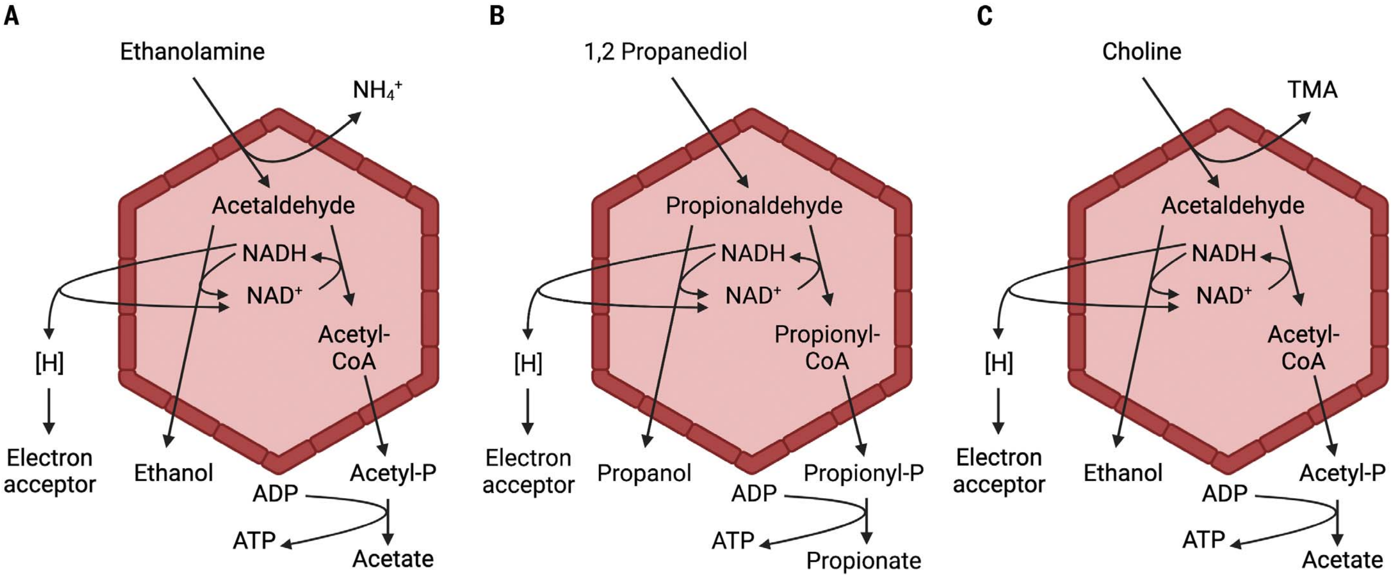


Fig. 4. Bacterial microcompartments that are powered by electron acceptors. (**A** to **C**) The catabolism of ethanolamine (A), 1,2 propanediol (B), and choline (C) requires bacterial microcompartments (hexagons) to shelter the cytosol from toxic intermediates (acetaldehyde or propionaldehyde). Catabolism through these microcompartments requires a respiratory electron acceptor, presumably because it enables bacteria to produce less alcohol (ethanol or propanol) and more acid (acetate or propionate), which increases the ATP yield (*122*). NADH and NAD^+, reduced and oxidized forms of nicotinamide adenine dinucleotide; ADP, adenosine diphosphate; NH_4^+, ammonium; [H], electron; CoA, coenzyme A; P, phosphate [created with BioRender.com].

risk for all-cause mortality in patients in a concentration-dependent manner (*131*). Thus, a rise in circulating levels of a metabolite linked to cardiovascular disease (i.e., TMAO) is triggered because multiple components of a high-energy, low-fiber diet act synergistically to change the environment of the gut microbiota (*96*). These findings illustrate that anaerobiosis not only balances the microbiota composition in the colon, but also maintains an environment that limits the generation of potentially harmful microbiota-derived metabolites, such as TMA.

Future directions

One limitation of relying on measuring microbiota composition and gene expression alone to define a balanced microbiome is that this microbe-centric approach does not consider how the host influences the environment inhabited by the microbiota. The term microbiome is derived from biome, which includes an environment and all the microbes within it (*132*). Therefore, a panel of international experts from the MicrobiomeSupport project recently proposed to define the microbiome as the microbiota and its environment (*4*). This viewpoint suggests that, in addition to measuring microbiota composition and gene expression, measurements of the host environment need to be performed to understand what constitutes a healthy microbiome.

The available evidence suggests that the host controls the environment of the microbiota using a similar hierarchy of redox reactions that also dictates the composition of microbial communities in abiotic environments (Fig. 1). By controlling the availability of oxygen and nitrate (Fig. 2), the host determines which redox reactions are available for generating ATP to fuel microbial growth. The presence of oxygen and nitrate in the small intestine results in a dominance of facultatively anaerobic bacteria because the use of these exogenous electron acceptors maximizes ATP production. The absence of oxygen and nitrate renders the redox reactions performed by obligately anaerobic primary fermenters the most successful metabolic strategy, resulting in their preeminence in the large intestine (Fig. 1). In addition to shaping the composition of the microbiota, limiting the availability of oxygen and nitrate in the colon also contributes to homeostasis by curbing metabolic pathways that produce potentially harmful microbiota-derived metabolites, such as TMA (Fig. 5). Mechanisms that control the availability of respiratory electron acceptors in the intestine could therefore be viewed as part of an immune system subdivision that shapes the microbiota to sustain host health (*15*, *18*, *19*).

Taking this perspective on host immunity implies that dysbiosis is characterized by an underlying impairment of host immune functions that regulate the microbiota and microbial metabolism (*10*, *19*) (Fig. 3). Measurements of these host functions could provide a more quantitative insight into the state of microbiome homeostasis than current microbe-centric approaches, which cannot be used to quantify gut homeostasis. For example, measurements of oxygen and nitrate concentrations along the longitudinal axis of the intestine could be used to determine a normal range that is characteristic of a healthy or normally functioning individual. This could perhaps be accomplished indirectly by monitoring a shift in epithelial energy metabolism using technologies such as intraluminal fluorescence lifetime imaging (*133*). Measurements within the normal range might indicate homeostasis, whereas values outside this range would indicate dysbiosis. Thus, determining the normal range of host functions that regulate the microbiota would provide units of measurement to quantify homeostasis and dysbiosis.

To be comprehensive, this approach would have to be expanded to include measurements of other host parameters relevant for microbial growth in the colon, including factors released in the upper gastrointestinal tract (*19*). For example, an impairment of digestive functions in the small intestine can have downstream effects on the fecal microbiota composition. Bile acids are absorbed to >95% in the ileum, but conditions that impair their absorption increase fecal bile concentrations (*134*). Deconjugation of bile by the colonic microbiota liberates taurine, which is used as an electron acceptor that is converted to sulfide (S^{2-}) by bacteria belonging to the class *Deltaproteobacteria* (*121*, *135*). By increasing taurine concentrations in the colon, impaired bile absorption therefore drives a bloom of *Deltaproteobacteria* (*121*). In turn, a bloom of *Deltaproteobacteria* exacerbates inflammation

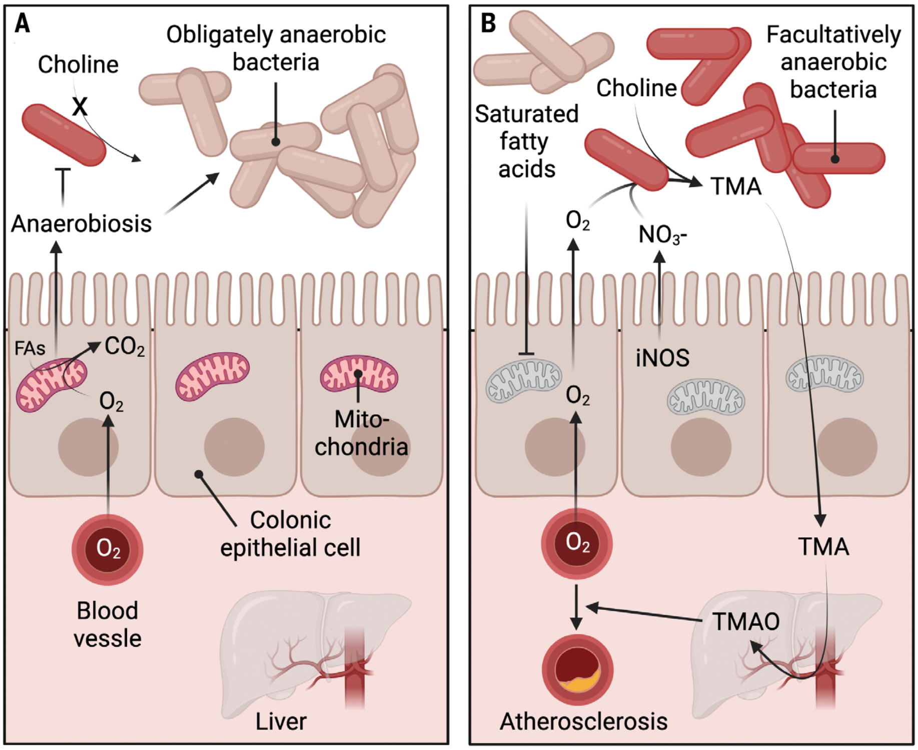


Fig. 5. Dysbiosis links diet to atherosclerosis. (**A**) During homeostasis, oxidation of fatty acids in the mitochondria results in high epithelial oxygen consumption, which maintains the epithelium in a state of physiological hypoxia. Epithelial hypoxia limits diffusion of oxygen into the colonic lumen, which preserves anaerobiosis. In turn, anaerobiosis maintains a dominance of obligately anaerobic primary fermenters (obligately anaerobic bacteria) within the colonic microbiota. The paucity of exogenous electron acceptors limits choline catabolism in microcompartments of facultatively anaerobic bacteria, which are minority species within the colonic microbiota during homeostasis. (**B**) Chronic high-fat intake increases epithelial oxygenation in the colon because saturated fatty acids decrease mitochondrial bioenergetics. In turn, increased levels of oxygen and nitrate in the colonic lumen trigger dysbiosis, which is characterized by an elevated abundance of facultatively anaerobic bacteria. Furthermore, oxygen and nitrate power choline catabolism in bacterial microcompartments, thereby escalating TMA production by the microbiota. TMA is absorbed by the host and converted to TMAO in the liver. The resulting increase in circulating TMAO levels is linked to the development of atherosclerosis [created with BioRender.com].

in a mouse model of inflammatory bowel disease (*128*). This example illustrates that quantity and composition of bile acid metabolites provide important measures for the state of colonic microbiome homeostasis. Another important host parameter shaping microbial growth in the colon is the integrity of the mucus barrier (*82*). In addition to the physical barrier formed by mucin glycans of the inner mucus layer, NOX1-derived reactive oxygen species form a chemical barrier that presumably keeps obligately anaerobic bacteria at a distance from the epithelial surface (*136*, *137*). There are likely additional host parameters that would need to be considered to be comprehensive, but eventually such an approach has the potential to generate diagnostic tools to assess the state of colonic microbiome homeostasis more quantitatively.

In addition to making it feasible to obtain some quantitative measure for the state of a microbiome, the concept of host regulation of the microbiota makes it possible to think about targeting host functions with therapeutic approaches to remediate dysbiosis (*19*). Dysbiosis in the colonic microbiota is often triggered because impaired mitochondrial function in the colonic epithelium disrupts epithelial hypoxia and disturbs anaerobiosis in the lumen (*14*, *17*). Recent studies have suggested that drugs that revive mitochondrial bioenergetics in the colonic epithelium can be used to restore gut homeostasis. An important positive regulator of mitochondrial bioenergetics in the colonic epithelium is the nuclear receptor peroxisome proliferator–activated receptor gamma (PPAR-γ) (*54*). Mitochondrial bioenergetics in the colonic epithelium can be strengthened by treatment with the PPAR-γ agonist 5-aminosalicylic acid [(5-ASA), the active ingredient of mesalamine] (*138*), which is poorly absorbed in the small intestine and acts specifically on epithelial cells in the colon (*111*). 5-ASA treatment normalizes the colonic microbiota composition by checking the growth of facultatively anaerobic bacteria in patients with mild to moderate ulcerative colitis (*139*), in mouse models of colitis (*92*, *111*), or in those fed a high-fat diet (*96*). Furthermore, strengthening mitochondrial bioenergetics using 5-ASA or resveratrol, a SIRTUIN 1 agonist (*140*, *141*), reduces circulating TMAO levels in mouse models of high-fat diet (*96*, *142*), which suggests that restoring epithelial hypoxia also limits production of potentially harmful metabolites by the colonic microbiota.

Measuring host homeostasis is still hampered by the need to define all of the factors used by the host to regulate its microbiota. Because gut dysbiosis is linked to a broad range of noncommunicable human diseases (*1*, *13*), consideration of host-driven microbiota regulation has the potential for guiding the development of remedies for a wide spectrum of illnesses, a prospect that makes the study of host factors governing the gut microbiota perhaps one of the most exciting areas in microbiome research.

REFERENCES AND NOTES

1. P. D. Cani, Gut microbiota – at the intersection of everything? *Nat. Rev. Gastroenterol. Hepatol.* **14**, 321–322 (2017). doi: 10.1038/nrgastro.2017.54; pmid: 28442782
2. C. Petersen, J. L. Round, Defining dysbiosis and its influence on host immunity and disease. *Cell. Microbiol.* **16**, 1024–1033 (2014). doi: 10.1111/cmi.12308; pmid: 24798552
3. J. E. McDonald, J. R. Marchesi, B. Koskella, Application of ecological and evolutionary theory to microbiome community dynamics across systems. *Proc. R. Soc. B.* **287**, 20202886 (2020). doi: 10.1098/rspb.2020.2886; pmid: 33352082
4. G. Berg *et al*., Microbiome definition re-visited: Old concepts and new challenges. *Microbiome* **8**, 103 (2020). doi: 10.1186/s40168-020-00875-0; pmid: 32605663
5. J. Tap *et al*., Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11**, 2574–2584 (2009). doi: 10.1111/j.1462-2920.2009.01982.x; pmid: 19601958
6. C. A. Lozupone, J. I. Stombaugh, J. I. Gordon, J. K. Jansson, R. Knight, Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012). doi: 10.1038/nature11550; pmid: 22972295
7. S. W. Olesen, E. J. Alm, Dysbiosis is not an answer. *Nat. Microbiol.* **1**, 16228 (2016). doi: 10.1038/nmicrobiol.2016.228; pmid: 27886190
8. E. Yong, "There Is No 'Healthy' Microbiome," *The New York Times*, 2 November 2014, p. 4.
9. L. Proctor, Priorities for the next 10 years of human microbiome research. *Nature* **569**, 623–625 (2019). doi: 10.1038/d41586-019-01654-0; pmid: 31142863
10. C. R. Tiffany, A. J. Bäumler, Dysbiosis: From fiction to function. *Am. J. Physiol. Gastrointest. Liver Physiol.* **317**, G602–G608 (2019). doi: 10.1152/ajpgi.00230.2019; pmid: 31509433
11. K. R. Foster, J. Schluter, K. Z. Coyte, S. Rakoff-Nahoum, The evolution of the host microbiome as an ecosystem on a leash. *Nature* **548**, 43–51 (2017). doi: 10.1038/nature23292; pmid: 28770836
12. M. X. Byndloss, S. R. Pernitzsch, A. J. Bäumler, Healthy hosts rule within: Ecological forces shaping the gut microbiota. *Mucosal Immunol.* **11**, 1299–1305 (2018). doi: 10.1038/s41385-018-0010-y; pmid: 29743614
13. M. X. Byndloss, A. J. Bäumler, The germ-organ theory of non-communicable diseases. *Nat. Rev. Microbiol.* **16**, 103–110 (2018). doi: 10.1038/nrmicro.2017.158; pmid: 29307890
14. Y. Litvak, M. X. Byndloss, A. J. Bäumler, Colonocyte metabolism shapes the gut microbiota. *Science* **362**, eaat9076 (2018). doi: 10.1126/science.aat9076; pmid: 30498100

15. Y. Litvak, A. J. Bäumler, Microbiota-Nourishing Immunity: A Guide to Understanding Our Microbial Self. *Immunity* **51**, 214–224 (2019). doi: 10.1016/j.immuni.2019.08.003; pmid: 31433969
16. A. C. Engevik, M. A. Engevik, Exploring the impact of intestinal ion transport on the gut microbiota. *Comput. Struct. Biotechnol. J.* **19**, 134–144 (2020). doi: 10.1016/j.csbj.2020.12.008; pmid: 33425246
17. Y. Litvak, M. X. Byndloss, R. M. Tsolis, A. J. Bäumler, Dysbiotic Proteobacteria expansion: A microbial signature of epithelial dysfunction. *Curr. Opin. Microbiol.* **39**, 1–6 (2017). doi: 10.1016/j.mib.2017.07.003; pmid: 28783509
18. M. X. Byndloss, Y. Litvak, A. J. Bäumler, Microbiota-nourishing Immunity and Its Relevance for Ulcerative Colitis. *Inflamm. Bowel Dis.* **25**, 811–815 (2019). doi: 10.1093/ibd/izz004; pmid: 30698700
19. B. M. Miller, A. J. Bäumler, The Habitat Filters of Microbiota-Nourishing Immunity. *Annu. Rev. Immunol.* **39**, 1–18 (2021). doi: 10.1146/annurev-immunol-101819-024945; pmid: 33902314
20. A. H. Stouthamer, A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek* **39**, 545–565 (1973). doi: 10.1007/BF02578899; pmid: 4148026
21. G. Wu, N. Zhao, C. Zhang, Y. Y. Lam, L. Zhao, Guild-based analysis for understanding gut microbiome in human health and diseases. *Genome Med.* **13**, 22 (2021). doi: 10.1186/s13073-021-00840-y; pmid: 33563315
22. A. J. Stams, C. M. Plugge, Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat. Rev. Microbiol.* **7**, 568–577 (2009). doi: 10.1038/nrmicro2166; pmid: 19609258
23. R. A. Schmitz, R. Daniel, U. Deppenmeier, G. Gottschalk, in *Prokaryotes: A Handbook on the Biology of Bacteria*, vol. 2, M. Dworkin, Ed. (Springer, ed. 3, 2006), pp. 86–101.
24. D. A. Stahl, M. Hullar, S. Davidson, in *Prokaryotes: A Handbook on the Biology of Bacteria*, vol. 1, M. Dworkin, Ed. (Springer, ed. 3, 2006), pp. 299–327.
25. K. H. Nealson, D. A. Stahl, in *Geomicrobiology: Interactions Between Microbes and Minerals*, vol. 35, J. F. Banfield, K. H. Nealson, Eds. (Mineralogical Society of America, 1997), pp. 5–34.
26. E. Puertollano, S. Kolida, P. Yaqoob, Biological significance of short-chain fatty acid metabolism by the intestinal microbiome. *Curr. Opin. Clin. Nutr. Metab. Care* **17**, 139–144 (2014). doi: 10.1097/MCO.0000000000000025; pmid: 24389673
27. E. S. Friedman *et al.*, Microbes vs. chemistry in the origin of the anaerobic gut lumen. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4170–4175 (2018). doi: 10.1073/pnas.1718635115; pmid: 29610310
28. M. Adeolu, S. Alnajar, S. Naushad, R. S. Gupta, Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': Proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **66**, 5575–5599 (2016). doi: 10.1099/ijsem.0.001485; pmid: 27620848
29. C. Tropini, K. A. Earle, K. C. Huang, J. L. Sonnenburg, The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host Microbe* **21**, 433–442 (2017). doi: 10.1016/j.chom.2017.03.010; pmid: 28407481
30. O. H. Sundin *et al.*, The human jejunum has an endogenous microbiota that differs from those in the oral cavity and colon. *BMC Microbiol.* **17**, 160 (2017). doi: 10.1186/s12866-017-1059-6; pmid: 28716079
31. R. Sender, S. Fuchs, R. Milo, Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biol.* **14**, e1002533 (2016). doi: 10.1371/journal.pbio.1002533; pmid: 27541692
32. M. Levy, E. Blacher, E. Elinav, Microbiome, metabolites and host immunity. *Curr. Opin. Microbiol.* **35**, 8–15 (2017). doi: 10.1016/j.mib.2016.10.003; pmid: 27883933
33. A. Koh, F. De Vadder, P. Kovatcheva-Datchary, F. Bäckhed, From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **165**, 1332–1345 (2016). doi: 10.1016/j.cell.2016.05.041; pmid: 27259147
34. P. Louis, G. L. Hold, H. J. Flint, The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* **12**, 661–672 (2014). doi: 10.1038/nrmicro3344; pmid: 25198138
35. J. L. McCarville, G. Y. Chen, V. D. Cuevas, K. Troha, J. S. Ayres, Microbiota Metabolites in Health and Disease. *Annu. Rev. Immunol.* **38**, 147–170 (2020). doi: 10.1146/annurev-immunol-071219-125715; pmid: 32340573
36. M. G. Rooks, W. S. Garrett, Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* **16**, 341–352 (2016). doi: 10.1038/nri.2016.42; pmid: 27231050
37. Y. Tong, T. Marion, G. Schett, Y. Luo, Y. Liu, Microbiota and metabolites in rheumatic diseases. *Autoimmun. Rev.* **19**, 102530 (2020). doi: 10.1016/j.autrev.2020.102530; pmid: 32240855
38. S. R. Gill *et al.*, Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006). doi: 10.1126/science.1124234; pmid: 16741115
39. J. van Duynhoven *et al.*, Metabolic fate of polyphenols in the human superorganism. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4531–4538 (2011). doi: 10.1073/pnas.1000098107; pmid: 20615997
40. J. Bures *et al.*, Small intestinal bacterial overgrowth syndrome. *World J. Gastroenterol.* **16**, 2978–2990 (2010). doi: 10.3748/wjg.v16.i24.2978; pmid: 20572300
41. I. Adlerberth, A. E. Wold, Establishment of the gut microbiota in Western infants. *Acta Paediatr.* **98**, 229–238 (2009). doi: 10.1111/j.1651-2227.2008.01060.x; pmid: 19143664
42. K. Orrhage, C. E. Nord, Factors controlling the bacterial colonization of the intestine in breastfed infants. *Acta Paediatr. Suppl.* **88**, 47–57 (1999). doi: 10.1111/j.1651-2227.1999.tb01300.x; pmid: 10569223
43. L. J. Mata, J. J. Urrutia, Intestinal Colonization of Breast-Fed Children in a Rural Area of Low Socioeconomic Level. *Ann. N.Y. Acad. Sci.* **176**, 93–109 (1971). doi: 10.1111/j.1749-6632.1971.tb34996.x
44. V. O. Rotimi, B. I. Duerden, The development of the bacterial flora in normal neonates. *J. Med. Microbiol.* **14**, 51–62 (1981). doi: 10.1099/00222615-14-1-51; pmid: 7463467
45. P. L. Stark, A. Lee, The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life. *J. Med. Microbiol.* **15**, 189–203 (1982). doi: 10.1099/00222615-15-2-189; pmid: 7143428
46. D. Halpern *et al.*, Do Primocolonizing Bacteria Enable *Bacteroides thetaiotaomicron* Intestinal Colonization Independently of the Capacity To Consume Oxygen? *mSphere* **6**, e00232-19 (2021). doi: 10.1128/mSphere.00232-19; pmid: 33952662
47. L. Zheng, C. J. Kelly, S. P. Colgan, Physiologic hypoxia and oxygen homeostasis in the healthy intestine. A Review in the Theme: Cellular Responses to Hypoxia. *Am. J. Physiol. Cell Physiol.* **309**, C350–C360 (2015). doi: 10.1152/ajpcell.00191.2015; pmid: 26179603
48. C. Matziouridou *et al.*, iNOS- and NOX1-dependent ROS production maintains bacterial homeostasis in the ileum of mice. *Mucosal Immunol.* **11**, 774–784 (2018). doi: 10.1038/mi.2017.106; pmid: 29210363
49. R. A. Hoffman *et al.*, Constitutive expression of inducible nitric oxide synthase in the mouse ileal mucosa. *Am. J. Physiol.* **272**, G383–G392 (1997). doi: 10.1152/ajpgi.1997.272.2.G383; pmid: 9124364
50. L. Zhu, C. Gunn, J. S. Beckman, Bactericidal activity of peroxynitrite. *Arch. Biochem. Biophys.* **298**, 452–457 (1992). doi: 10.1016/0003-9861(92)90434-X; pmid: 1416976
51. M. A. De Groote *et al.*, Genetic and redox determinants of nitric oxide cytotoxicity in a Salmonella typhimurium model. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6399–6403 (1995). doi: 10.1073/pnas.92.14.6399; pmid: 7604003
52. C. Szabó, H. Ischiropoulos, R. Radi, Peroxynitrite: Biochemistry, pathophysiology and development of therapeutics. *Nat. Rev. Drug Discov.* **6**, 662–680 (2007). doi: 10.1038/nrd2222; pmid: 17667957
53. F. Rivera-Chávez *et al.*, Energy Taxis toward Host-Derived Nitrate Supports a *Salmonella* Pathogenicity Island 1-Independent Mechanism of Invasion. *mBio* **7**, e00960-16 (2016). doi: 10.1128/mBio.00960-16; pmid: 27435462
54. M. X. Byndloss *et al.*, Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**, 570–575 (2017). doi: 10.1126/science.aam9949; pmid: 28798125
55. Human Microbiome Project Consortium, Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012). doi: 10.1038/nature11234; pmid: 22699609
56. N. M. Koropatkin, E. A. Cameron, E. C. Martens, How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012). doi: 10.1038/nrmicro2746; pmid: 22491358
57. K. A. Gwee, Fiber, FODMAPs, flora, flatulence, and the functional bowel disorders. *J. Gastroenterol. Hepatol.* **25**, 1335–1336 (2010). doi: 10.1111/j.1440-1746.2010.06433.x; pmid: 20659218
58. D. Rothschild *et al.*, Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018). doi: 10.1038/nature25973; pmid: 29489753
59. P. Vangay *et al.*, US Immigration Westernizes the Human Gut Microbiome. *Cell* **175**, 962–972.e10 (2018). doi: 10.1016/j.cell.2018.10.029; pmid: 30388453
60. K. Stagaman *et al.*, Market Integration Predicts Human Gut Microbiome Attributes across a Gradient of Economic Development. *mSystems* **3**, e00122-17 (2018). doi: 10.1128/mSystems.00122-17; pmid: 29507896
61. N. Kirmiz, R. C. Robinson, I. M. Shah, D. Barile, D. A. Mills, Milk Glycans and Their Interaction with the Infant-Gut Microbiota. *Annu. Rev. Food Sci. Technol.* **9**, 429–450 (2018). doi: 10.1146/annurev-food-030216-030207; pmid: 29580136
62. D. A. Sela, D. A. Mills, Nursing our microbiota: Molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol.* **18**, 298–307 (2010). doi: 10.1016/j.tim.2010.03.008; pmid: 20409714
63. J. B. German, S. L. Freeman, C. B. Lebrilla, D. A. Mills, Human milk oligosaccharides: Evolution, structures and bioselectivity as substrates for intestinal bacteria. *Nestle Nutr. Workshop Ser. Pediatr. Program.* **62**, 205–222 (2008). doi: 10.1159/000146322; pmid: 18626202
64. D. A. Sela *et al.*, The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18964–18969 (2008). doi: 10.1073/pnas.0809584105; pmid: 19033196
65. D. A. Sela *et al.*, *Bifidobacterium longum* subsp. *infantis* ATCC 15697 α-fucosidases are active on fucosylated human milk oligosaccharides. *Appl. Environ. Microbiol.* **78**, 795–803 (2012). doi: 10.1128/AEM.06762-11; pmid: 22138995
66. L. C. Roger, A. Costabile, D. T. Holland, L. Hoyles, A. L. McCartney, Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* **156**, 3329–3341 (2010). doi: 10.1099/mic.0.043224-0; pmid: 20864478
67. R. I. Mackie, A. Sghir, H. R. Gaskins, Developmental microbial ecology of the neonatal gastrointestinal tract. *Am. J. Clin. Nutr.* **69**, 1035s–1045s (1999). doi: 10.1093/ajcn/69.5.1035s; pmid: 10232646
68. A. El Kaoutari, F. Armougom, J. I. Gordon, D. Raoult, B. Henrissat, The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11**, 497–504 (2013). doi: 10.1038/nrmicro3050; pmid: 23748339
69. C. R. Tiffany *et al.*, The metabolic footprint of *Clostridia* and *Erysipelotrichia* reveals their role in depleting sugar alcohols in the cecum. *Microbiome* **9**, 174 (2021). doi: 10.1186/s40168-021-01123-9; pmid: 34412707
70. J. E. Gordon, Weanling Diarrhea: A Synergism of Nutrition and Infection. *Nutr. Rev.* **22**, 161–163 (1964). doi: 10.1111/j.1753-4887.1964.tb07772.x; pmid: 14187680
71. A. W. L. Rogers, R. M. Tsolis, A. J. Bäumler, *Salmonella* versus the Microbiome. *Microbiol. Mol. Biol. Rev.* **85**, e00027-19 (2020). doi: 10.1128/MMBR.00027-19; pmid: 33361269
72. P. G. Kopelman, Obesity as a medical problem. *Nature* **404**, 635–643 (2000). doi: 10.1038/35007508; pmid: 10766250
73. B. A. Swinburn *et al.*, The global obesity pandemic: Shaped by global drivers and local environments. *Lancet* **378**, 804–814 (2011). doi: 10.1016/S0140-6736(11)60813-1; pmid: 21872749
74. A. Hruby, F. B. Hu, The Epidemiology of Obesity: A Big Picture. *PharmacoEconomics* **33**, 673–689 (2015). doi: 10.1007/s40273-014-0243-x; pmid: 25471927
75. E. Amabebe, F. O. Robert, T. Agbalalah, E. S. F. Orubu, Microbial dysbiosis-induced obesity: Role of gut microbiota in homoeostasis of energy metabolism. *Br. J. Nutr.* **123**, 1127–1137 (2020). doi: 10.1017/S0007114520000380; pmid: 32008579
76. Z. Wang *et al.*, Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011). doi: 10.1038/nature09922; pmid: 21475195
77. W. S. Garrett *et al.*, *Enterobacteriaceae* act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292–300 (2010). doi: 10.1016/j.chom.2010.08.004; pmid: 20833380
78. M. Song, W. S. Garrett, A. T. Chan, Nutrients, foods, and colorectal cancer prevention. *Gastroenterology* **148**, 1244–1260.e16 (2015). doi: 10.1053/j.gastro.2014.12.035; pmid: 25575572

79. J. B. Lynch, J. L. Sonnenburg, Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a Bacteroides hybrid two-component system. *Mol. Microbiol.* **85**, 478–491 (2012). doi: 10.1111/j.1365-2958.2012.08123.x; pmid: 22686399
80. J. L. Sonnenburg *et al*., Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005). doi: 10.1126/science.1109051; pmid: 15790854
81. M. S. Desai *et al*., A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339–1353.e21 (2016). doi: 10.1016/j.cell.2016.10.043; pmid: 27863247
82. M. E. Johansson *et al*., The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15064–15069 (2008). doi: 10.1073/pnas.0803124105; pmid: 18806221
83. M. E. Johansson *et al*., Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLOS ONE* **5**, e12238 (2010). doi: 10.1371/journal.pone.0012238; pmid: 20805871
84. S. van der Post *et al*., Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. *Gut* **68**, 2142–2151 (2019). doi: 10.1136/gutjnl-2018-317571; pmid: 30914450
85. M. E. Johansson *et al*., Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* **63**, 281–291 (2014). doi: 10.1136/gutjnl-2012-303207; pmid: 23426893
86. C. Jang *et al*., The Small Intestine Converts Dietary Fructose into Glucose and Organic Acids. *Cell Metab.* **27**, 351–361.e3 (2018). doi: 10.1016/j.cmet.2017.12.016; pmid: 29414685
87. P. J. Turnbaugh, F. Bäckhed, L. Fulton, J. I. Gordon, Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008). doi: 10.1016/j.chom.2008.02.015; pmid: 18407065
88. T. Sen *et al*., Diet-driven microbiota dysbiosis is associated with vagal remodeling and obesity. *Physiol. Behav.* **173**, 305–317 (2017). doi: 10.1016/j.physbeh.2017.02.027; pmid: 28249783
89. V. Volynets *et al*., Intestinal Barrier Function and the Gut Microbiome Are Differentially Affected in Mice Fed a Western-Style Diet or Drinking Water Supplemented with Fructose. *J. Nutr.* **147**, 770–780 (2017). doi: 10.3945/jn.116.242859; pmid: 28356436
90. S. Khan *et al*., Dietary simple sugars alter microbial ecology in the gut and promote colitis in mice. *Sci. Transl. Med.* **12**, eaay6218 (2020). doi: 10.1126/scitranslmed.aay6218; pmid: 33115951
91. M. Gulhane *et al*., High Fat Diets Induce Colonic Epithelial Cell Stress and Inflammation that is Reversed by IL-22. *Sci. Rep.* **6**, 28990 (2016). doi: 10.1038/srep28990; pmid: 27350069
92. J. Y. Lee *et al*., High-Fat Diet and Antibiotics Cooperatively Impair Mitochondrial Bioenergetics to Trigger Dysbiosis that Exacerbates Pre-inflammatory Bowel Disease. *Cell Host Microbe* **28**, 273–284.e6 (2020). doi: 10.1016/j.chom.2020.06.001; pmid: 32668218
93. D. Arnone *et al*., Long-Term Overconsumption of Fat and Sugar Causes a Partially Reversible Pre-inflammatory Bowel Disease State. *Front. Nutr.* **8**, 758518 (2021). doi: 10.3389/fnut.2021.758518; pmid: 34869528
94. A. R. Cardoso, P. A. Kakimoto, A. J. Kowaltowski, Diet-sensitive sources of reactive oxygen species in liver mitochondria: Role of very long chain acyl-CoA dehydrogenases. *PLOS ONE* **8**, e77088 (2013). doi: 10.1371/journal.pone.0077088; pmid: 24116206
95. P. A. Kakimoto, F. K. Tamaki, A. R. Cardoso, S. R. Marana, A. J. Kowaltowski, H_2O_2 release from the very long chain acyl-CoA dehydrogenase. *Redox Biol.* **4**, 375–380 (2015). doi: 10.1016/j.redox.2015.02.003; pmid: 25728796
96. W. Yoo *et al*., High-fat diet-induced colonocyte dysfunction escalates microbiota-derived trimethylamine *N*-oxide. *Science* **373**, 813–818 (2021). doi: 10.1126/science.aba3683; pmid: 34385401
97. S. E. Winter, C. A. Lopez, A. J. Bäumler, The dynamics of gut-associated microbial communities during inflammation. *EMBO Rep.* **14**, 319–327 (2013). doi: 10.1038/embor.2013.27; pmid: 23478337
98. F. Rivera-Chávez, C. A. Lopez, A. J. Bäumler, Oxygen as a driver of gut dysbiosis. *Free Radic. Biol. Med.* **105**, 93–101 (2017). doi: 10.1016/j.freeradbiomed.2016.09.022; pmid: 27677568
99. M. A. Henson, P. Phalak, Microbiota dysbiosis in inflammatory bowel diseases: In silico investigation of the oxygen hypothesis. *BMC Syst. Biol.* **11**, 145 (2017). doi: 10.1186/s12918-017-0522-1; pmid: 29282051
100. A. M. Spees *et al*., Streptomycin-induced inflammation enhances *Escherichia coli* gut colonization through nitrate respiration. *mBio* **4**, e00430-13 (2013). doi: 10.1128/mBio.00430-13; pmid: 23820397
101. C. A. Lopez *et al*., Phage-mediated acquisition of a type III secreted effector protein boosts growth of *Salmonella* by nitrate respiration. *mBio* **3**, e00143-12 (2012). doi: 10.1128/mBio.00143-12; pmid: 22691391
102. C. A. Lopez, F. Rivera-Chávez, M. X. Byndloss, A. J. Bäumler, The Periplasmic Nitrate Reductase NapABC Supports Luminal Growth of *Salmonella enterica* Serovar Typhimurium during Colitis. *Infect. Immun.* **83**, 3470–3478 (2015). doi: 10.1128/IAI.00351-15; pmid: 26099579
103. P. A. McLaughlin *et al*., Inflammatory monocytes provide a niche for *Salmonella* expansion in the lumen of the inflamed intestine. *PLOS Pathog.* **15**, e1007847 (2019). doi: 10.1371/journal.ppat.1007847; pmid: 31306468
104. S. Wang *et al*., Infection-Induced Intestinal Dysbiosis Is Mediated by Macrophage Activation and Nitrate Production. *mBio* **10**, e00935-19 (2019). doi: 10.1128/mBio.00935-19; pmid: 31138751
105. C. A. Lopez *et al*., Virulence factors enhance *Citrobacter rodentium* expansion through aerobic respiration. *Science* **353**, 1249–1253 (2016). doi: 10.1126/science.aag3042; pmid: 27634526
106. F. Rivera-Chávez *et al*., Depletion of Butyrate-Producing *Clostridia* from the Gut Microbiota Drives an Aerobic Luminal Expansion of *Salmonella*. *Cell Host Microbe* **19**, 443–454 (2016). doi: 10.1016/j.chom.2016.03.004; pmid: 27078066
107. C. C. Gillis *et al*., Dysbiosis-Associated Change in Host Metabolism Generates Lactate to Support *Salmonella* Growth. *Cell Host Microbe* **23**, 54–64.e6 (2018). doi: 10.1016/j.chom.2017.11.006; pmid: 29276172
108. Y. Litvak *et al*., Commensal *Enterobacteriaceae* Protect against *Salmonella* Colonization through Oxygen Competition. *Cell Host Microbe* **25**, 128–139.e5 (2019). doi: 10.1016/j.chom.2018.12.003; pmid: 30629913
109. S. E. Winter *et al*., Host-derived nitrate boosts growth of *E. coli* in the inflamed gut. *Science* **339**, 708–711 (2013). doi: 10.1126/science.1232467; pmid: 23393266
110. W. Zhu *et al*., Precision editing of the gut microbiota ameliorates colitis. *Nature* **553**, 208–211 (2018). doi: 10.1038/nature25172; pmid: 29323293
111. S. A. Cevallos *et al*., 5-Aminosalicylic Acid Ameliorates Colitis and Checks Dysbiotic *Escherichia coli* Expansion by Activating PPAR-γ Signaling in the Intestinal Epithelium. *mBio* **12**, e03227-20 (2021). doi: 10.1128/mBio.03227-20; pmid: 33468700
112. E. R. Hughes *et al*., Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host Microbe* **21**, 208–219 (2017). doi: 10.1016/j.chom.2017.01.005; pmid: 28182951
113. S. A. Pötgens *et al*., *Klebsiella oxytoca* expands in cancer cachexia and acts as a gut pathobiont contributing to intestinal dysfunction. *Sci. Rep.* **8**, 12321 (2018). doi: 10.1038/s41598-018-30569-5; pmid: 30120320
114. W. Zhu *et al*., Editing of the gut microbiota reduces carcinogenesis in mouse models of colitis-associated colorectal cancer. *J. Exp. Med.* **216**, 2378–2393 (2019). doi: 10.1084/jem.20181939; pmid: 31358565
115. S. A. Cevallos *et al*., Increased Epithelial Oxygenation Links Colitis to an Expansion of Tumorigenic Bacteria. *mBio* **10**, e02244-19 (2019). doi: 10.1128/mBio.02244-19; pmid: 31575772
116. N. R. Shin, T. W. Whon, J. W. Bae, *Proteobacteria*: Microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* **33**, 496–503 (2015). doi: 10.1016/j.tibtech.2015.06.011; pmid: 26210164
117. C. A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018). doi: 10.1038/nrmicro.2018.10; pmid: 29503457
118. M. Price-Carter, J. Tingey, T. A. Bobik, J. R. Roth, The alternative electron acceptor tetrathionate supports B_{12}-dependent anaerobic growth of *Salmonella enterica* serovar typhimurium on ethanolamine or 1,2-propanediol. *J. Bacteriol.* **183**, 2463–2475 (2001). doi: 10.1128/JB.183.8.2463-2475.2001; pmid: 11274105
119. P. Thiennimitr *et al*., Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17480–17485 (2011). doi: 10.1073/pnas.1107857108; pmid: 21969563
120. F. Faber *et al*., Respiration of Microbiota-Derived 1,2-propanediol Drives *Salmonella* Expansion during Colitis. *PLOS Pathog.* **13**, e1006129 (2017). doi: 10.1371/journal.ppat.1006129; pmid: 28056091
121. A. Stacy *et al*., Infection trains the host for microbiota-enhanced resistance to pathogens. *Cell* **184**, 615–627.e17 (2021). doi: 10.1016/j.cell.2020.12.011; pmid: 33453153
122. Z. Zeng *et al*., Bacterial Microcompartments Coupled with Extracellular Electron Transfer Drive the Anaerobic Utilization of Ethanolamine in *Listeria monocytogenes*. *mSystems* **6**, e01349-20 (2021). doi: 10.1128/mSystems.01349-20; pmid: 33850044
123. M. Viladomiu *et al*., Adherent-invasive *E. coli* metabolism of propanediol in Crohn's disease regulates phagocytes to drive intestinal inflammation. *Cell Host Microbe* **29**, 607–619.e8 (2021). doi: 10.1016/j.chom.2021.01.002; pmid: 33539767
124. A. Martínez-del Campo *et al*., Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *mBio* **6**, e00042-15 (2015). doi: 10.1128/mBio.00042-15; pmid: 25873372
125. Y. Zhu *et al*., Carnitine metabolism to trimethylamine by an unusual Rieske-type oxygenase from human microbiota. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4268–4273 (2014). doi: 10.1073/pnas.1316569111; pmid: 24591617
126. M. Martinez-Medina *et al*., Western diet induces dysbiosis with increased *E coli* in CEABAC10 mice, alters host barrier function favouring AIEC colonisation. *Gut* **63**, 116–124 (2014). doi: 10.1136/gutjnl-2012-304119; pmid: 23598352
127. N. Fei, L. Zhao, An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *ISME J.* **7**, 880–884 (2013). doi: 10.1038/ismej.2012.153; pmid: 23235292
128. S. Devkota *et al*., Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10*$^{-/-}$ mice. *Nature* **487**, 104–108 (2012). doi: 10.1038/nature11225; pmid: 22722865
129. W. H. Tang *et al*., Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584 (2013). doi: 10.1056/NEJMoa1109400; pmid: 23614584
130. R. A. Koeth *et al*., Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013). doi: 10.1038/nm.3145; pmid: 23563705
131. G. G. Schiattarella *et al*., Gut microbe-generated metabolite trimethylamine-N-oxide as cardiovascular risk biomarker: A systematic review and dose-response meta-analysis. *Eur. Heart J.* **38**, 2948–2956 (2017). doi: 10.1093/eurheartj/ehx342; pmid: 29020409
132. L. Tipton, J. L. Darcy, N. A. Hynson, A Developing Symbiosis: Enabling Cross-Talk Between Ecologists and Microbiome Scientists. *Front. Microbiol.* **10**, 292 (2019). doi: 10.3389/fmicb.2019.00292; pmid: 30842763
133. A. Alfonso-Garcia *et al*., Assessment of Murine Colon Inflammation Using Intraluminal Fluorescence Lifetime Imaging. *Molecules* **27**, 1317 (2022). doi: 10.3390/molecules27041317; pmid: 35209104
134. A. L. Ticho, P. Malhotra, P. K. Dudeja, R. K. Gill, W. A. Alrefai, in *Comprehensive Physiology*, Y. Prakash, Ed. (Wiley, 2020), pp. 21–56.
135. S. C. Peck *et al*., A glycyl radical enzyme enables hydrogen sulfide production by the human intestinal bacterium *Bilophila wadsworthia*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 3171–3176 (2019). doi: 10.1073/pnas.1815661116; pmid: 30718429
136. B. M. Miller *et al*., Anaerobic Respiration of NOX1-Derived Hydrogen Peroxide Licenses Bacterial Growth at the Colonic Surface. *Cell Host Microbe* **28**, 789–797.e5 (2020). doi: 10.1016/j.chom.2020.10.009; pmid: 33301718
137. R. B. Chanin *et al*., Epithelial-Derived Reactive Oxygen Species Enable AppBCX-Mediated Aerobic Respiration of *Escherichia coli* during Intestinal Inflammation. *Cell Host Microbe* **28**, 780–788.e5 (2020). doi: 10.1016/j.chom.2020.09.005; pmid: 33053375
138. C. Rousseaux *et al*., Intestinal antiinflammatory effect of 5-aminosalicylic acid is dependent on peroxisome proliferator–activated receptor-γ. *J. Exp. Med.* **201**, 1205–1215 (2005). doi: 10.1084/jem.20041948; pmid: 15824083

139. J. Xu *et al*., 5-Aminosalicylic Acid Alters the Gut Bacterial Microbiota in Patients With Ulcerative Colitis. *Front. Microbiol.* **9**, 1274 (2018). doi: 10.3389/fmicb.2018.01274; pmid: 29951050
140. H. Nakayama, T. Yaguchi, S. Yoshiya, T. Nishizaki, Resveratrol induces apoptosis MH7A human rheumatoid arthritis synovial cells in a sirtuin 1-dependent manner. *Rheumatol. Int.* **32**, 151–157 (2012). doi: 10.1007/s00296-010-1598-8; pmid: 20697895
141. K. Aquilano *et al*., Peroxisome proliferator-activated receptor γ co-activator 1α (PGC-1α) and sirtuin 1 (SIRT1) reside in mitochondria: Possible direct function in mitochondrial biogenesis. *J. Biol. Chem.* **285**, 21590–21599 (2010). doi: 10.1074/jbc.M109.070169; pmid: 20448046
142. M. L. Chen *et al*., Resveratrol Attenuates Trimethylamine-N-Oxide (TMAO)-Induced Atherosclerosis by Regulating TMAO Synthesis and Bile Acid Metabolism via Remodeling of the Gut Microbiota. *mBio* **7**, e02210-15 (2016). doi: 10.1128/mBio.02210-15; pmid: 27048804

ACKNOWLEDGMENTS

Funding: Work in R.M.T.'s laboratory was supported by NIH awards AI143253, AI149632, AI112949, AI089078, and AI109799. Work in A.J.B.'s laboratory was supported by Crohn's and Colitis Foundation of America award 650976 and by Public Health Service grants AI044170, AI096528, AI112445, AI112949, AI146432, and AI153069. **Competing interests:** The authors do not declare any conflicts of interest. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 10 March 2022; accepted 24 May 2022
10.1126/science.abp9960

REPORT

PALEOCLIMATE

Cenozoic evolution of deep ocean temperature from clumped isotope thermometry

A. N. Meckler[1]*, P. F. Sexton[2], A. M. Piasecki[1]†, T. J. Leutert[1], J. Marquardt[1], M. Ziegler[3], T. Agterhuis[3], L. J. Lourens[3], J. W. B. Rae[4], J. Barnet[4], A. Tripati[5], S. M. Bernasconi[6]

Characterizing past climate states is crucial for understanding the future consequences of ongoing greenhouse gas emissions. Here, we revisit the benchmark time series for deep ocean temperature across the past 65 million years using clumped isotope thermometry. Our temperature estimates from the deep Atlantic Ocean are overall much warmer compared with oxygen isotope–based reconstructions, highlighting the likely influence of changes in deep ocean pH and/or seawater oxygen isotope composition on classical oxygen isotope records of the Cenozoic. In addition, our data reveal previously unrecognized large swings in deep ocean temperature during early Eocene acute greenhouse warmth. Our results call for a reassessment of the Cenozoic history of ocean temperatures to achieve a more accurate understanding of the nature of climatic responses to tectonic events and variable greenhouse forcing.

Throughout the Cenozoic era (the past 65 million years), Earth has experienced a large range of climate states (*1*, *2*), moving from greenhouse climates with high levels of atmospheric CO_2 and minimal ice on land to an icehouse world with large-scale, bipolar glaciation. Reconstructions of past temperatures offer unique insights into the response of the climate system to varying levels of CO_2 and other influences such as tectonically forced changes in ocean circulation. These reconstructions additionally serve as crucial benchmarks for assessing the performance of climate models (*3*). Deep ocean temperature is commonly used to obtain globally representative temperature estimates on million-year time scales [e.g., (*4, 5*)] because the deep ocean constitutes an extremely large and comparatively stable heat reservoir.

The key approach to reconstructing deep ocean temperatures, especially through the early Cenozoic, has been the analysis of the oxygen isotopic composition of shells of benthic foraminifera ($\delta^{18}O_b$) from deep ocean sediments. $\delta^{18}O_b$ reflects deep ocean temperature but also seawater isotopic composition ($\delta^{18}O_{sw}$) at the time of shell formation. The canonical view has been that before Antarctica was continually glaciated, $\delta^{18}O_{sw}$ had an assumed "ice-free Earth" value of –0.9 to –1.2‰ (*1*, *6*), and that temperature was the single most important influence on $\delta^{18}O_b$. Paleocene and early to middle Eocene temperatures are therefore typically calculated directly from $\delta^{18}O_b$ using established temperature calibrations [e.g., (*7*)]. This approach has yielded deep ocean temperatures of up to 12°C (*1*) to 14°C (*6*) during the warmest part of the Cenozoic, the early Eocene climatic optimum [EECO, 52 to 49 million years ago (Ma)]. However, the long-standing assumptions made as part of this established approach are largely uncorroborated (*6*, *8*). $\delta^{18}O_{sw}$ indeed reflects continental storage of ice, but also that of groundwater, and can in addition be related to changes in salinity and influenced by long-term changes in interactions between seawater and oceanic crust. Furthermore, nonthermal, physiological influences on $\delta^{18}O_b$ are not well understood (*7*) but are evident from $\delta^{18}O_b$ offsets between different species [e.g., (*9*)].

Additional proxy constraints have previously been sought from the Mg/Ca ratio of benthic foraminifera [e.g., (*10, 11*)], a temperature proxy that is independent of $\delta^{18}O_{sw}$. However, like $\delta^{18}O_b$, the Mg/Ca proxy suffers from nonthermal influences such as changes in seawater Mg/Ca ratio and carbonate chemistry (*12*) and physiological effects necessitating species-specific calibrations (*8*). Given these potential confounding influences on previous proxy records, the Cenozoic evolution of deep ocean temperature remains highly uncertain.

Here, we report the first Cenozoic record of deep ocean temperature based on carbonate clumped isotope (Δ_{47}) thermometry. The clumped isotope method takes advantage of the temperature dependence of isotopic ordering within molecules (*13*). It has the benefit that it does not require knowledge about seawater composition and is not measurably affected by physiological factors during the formation of foraminifera shells (see the supplementary materials and methods). Our record is predominantly from sites with exceptional foraminifera preservation in the North Atlantic, a region of major influence on global ocean circulation and climate. The reported clumped isotope temperatures each combine 13 to 45 replicate analyses of different foraminifera species. Species- or genus-specific $\delta^{18}O_b$ and carbon isotopic composition ($\delta^{13}C$) are obtained from the same measurements. The Δ_{47} temperatures show many similarities, but also marked differences, compared with previous reconstructions of deep ocean temperatures based on $\delta^{18}O_b$ (Fig. 1A and fig. S6). We observed the same overall trend across the Cenozoic, attesting to the fact that the large-scale features of the canonical $\delta^{18}O_b$ dataset do indeed reflect overall cooling through this era. This confirmation is especially important in light of a recent controversial suggestion that attributes the Cenozoic trend of increasing $\delta^{18}O_b$ to post-depositional alteration during burial (*14*, *15*). However, Δ_{47} temperatures are considerably warmer than previous estimates during most of the Cenozoic (Fig. 1A), with deviations of 3° to 5°C in the middle Oligocene and late Eocene and 6° to 8°C in some late Paleocene and early Eocene intervals. Our Δ_{47} data also reveal large temperature changes in the early Eocene (48 to 56 Ma) that are not apparent in the $\delta^{18}O_b$-based record.

Most of our Δ_{47} data are from the North Atlantic, a region that is seldom represented in Cenozoic climate records. The divergences from $\delta^{18}O_b$-based reconstructions could thus reflect regional differences. However, our North Atlantic $\delta^{18}O_b$ and $\delta^{13}C$ data obtained alongside the Δ_{47} measurements generally show close agreement with equivalent global records (Fig. 1, B and C, and supplementary materials). We furthermore found excellent agreement of our Δ_{47} temperatures with previous and new Δ_{47}-based temperature estimates from other locations. Our middle Miocene temperature estimates agree with results from the Indian and Southern Ocean at that time (*16*, *17*) (Fig. 1A, purple triangles). For the Eocene, North Atlantic temperatures are similar to those in the tropical and South Atlantic Ocean [Fig. 1A, yellow circle and pink stars (*18,19*), including the South Atlantic location that represents the backbone for the early Cenozoic global $\delta^{18}O_b$ record. Although further validation of the globally representative nature of North Atlantic temperatures awaits detailed Δ_{47} records from multiple sites, the available data point to our Δ_{47} temperatures being representative of substantial parts of the global ocean.

The consistency in Δ_{47} temperatures between different locations furthermore argues against

[1]Bjerknes Centre for Climate Research and Department of Earth Science, University of Bergen, Bergen, Norway. [2]School of Environment, Earth and Ecosystem Sciences, The Open University, Milton Keynes, UK. [3]Faculty of Geosciences, Utrecht University, Utrecht, Netherlands. [4]School of Earth and Environmental Sciences, University of St. Andrews, St. Andrews, UK. [5]Department of Earth, Planetary, and Space Science, Department of Atmospheric and Oceanic Science, Institute of the Environment and Sustainability, American Indian Studies Center, Center for Diverse Leadership in Science, University of California, Los Angeles, Los Angeles, USA. [6]Department of Earth Science, ETH Zürich, Zürich, Switzerland.
***Corresponding author. Email: nele.meckler@uib.no**
†Present address: CIRES, University of Colorado, Boulder, CO, USA.

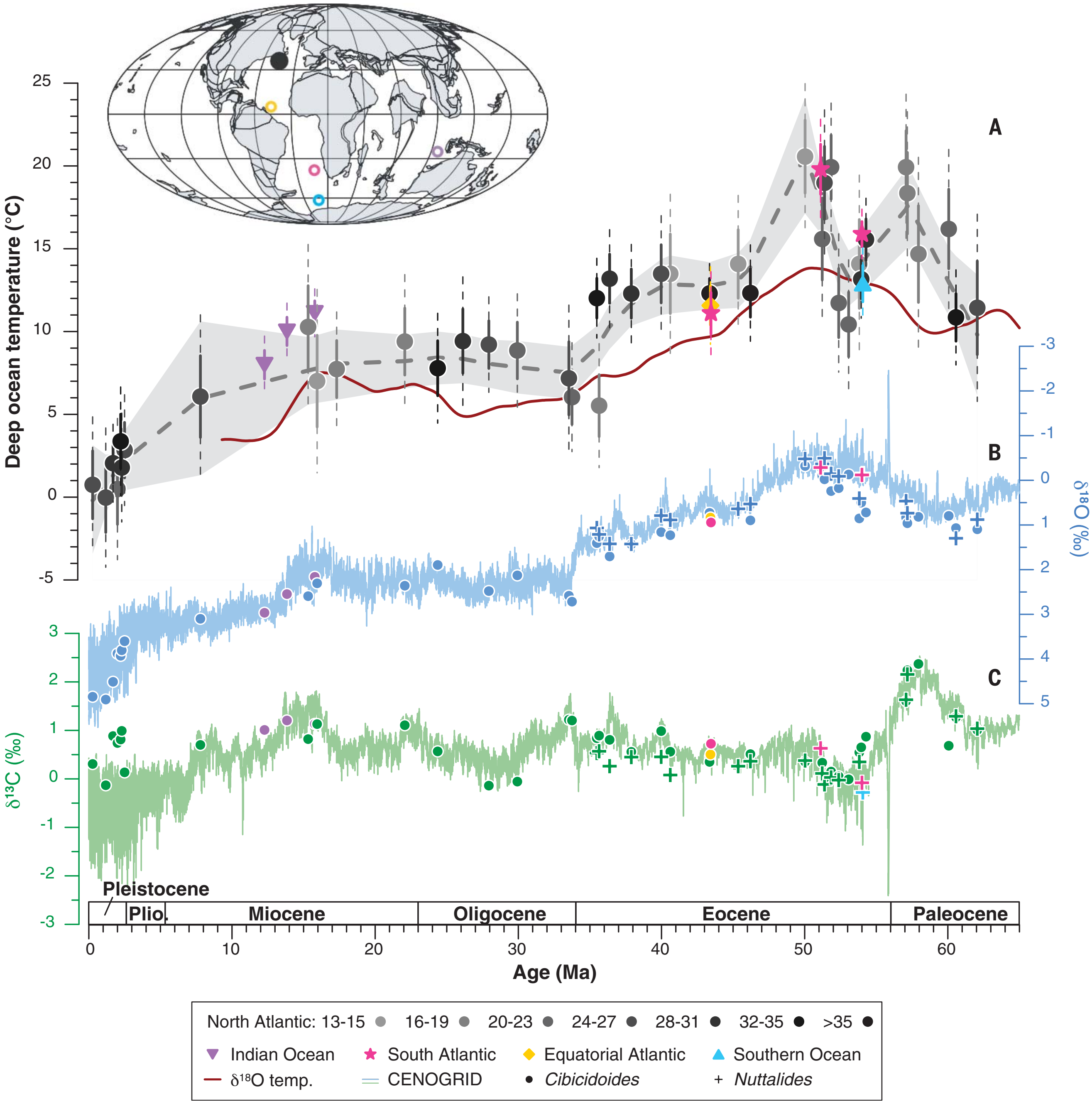


Fig. 1. Clumped isotope–based reconstruction of deep ocean temperatures compared with classical foraminifera isotope records. (**A**) Δ_{47} temperatures from the North Atlantic (gray symbols) compared with estimates from other sites (colored symbols; see inset map) in the Indian Ocean [middle Miocene (*16*)], equatorial and South Atlantic [this study (*18*, *19*)], and the Atlantic sector of the Southern Ocean (*18*). The early Eocene data from (*18*) span a time interval of ~200,000 years and were combined into one data point for each location. Error bars for Δ_{47} temperatures are 68% (solid lines) and 95% (stippled lines) confidence intervals. For the North Atlantic data, darkness of shading reflects robustness based on degree of replication, with ranges of replicate numbers specified in the figure legend. The other data points are based on 51 to >200 measurements (see data S1) reflected in the size of the error bars. The gray stippled line is a LOESS fit through the North Atlantic data with a Monte Carlo–based 95% confidence interval (see the supplementary materials), and is meant as visual guidance rather than as an alternative representation of the data. The red line is a previous estimate of deep ocean temperature based on $\delta^{18}O_b$ and sea-level reconstructions (*6*). (**B**) CENOGRID $\delta^{18}O_b$ record (*2*) and $\delta^{18}O_b$ results from species of the genus *Cibicidoides* (circles) and *Nuttalides* (crosses) from the clumped isotope analyses, corrected for interspecies offsets (*Cibicidoides:* +0.64‰; *Nuttalides:* +0.4‰). (**C**) CENOGRID $\delta^{13}C$ record (*2*) with results from *Cibicidoides* and *Nuttalides* species from the clumped isotope analyses. Note that the global CENOGRID record also includes corrections for offsets between sites and ocean basins, which we do not apply to our data. Typical offsets in $\delta^{18}O_b$ between Atlantic and Pacific sites (the latter interpreted as equivalent to global) are on the order of −0.2 to −0.3‰ for the Eocene to Oligocene (*2*), compared with 0 to −0.5‰ offsets in our *Nuttalides* data. Inset map shows site locations on a paleogeographic reconstruction for ~50 Ma (www.odsn.de/odsn/services/paleomap/paleomap.html). Also see fig. S4 for close-up views of early and late Cenozoic interval data.

a strong influence of postdepositional alteration of the Δ_{47} signal that would likely be site specific. Preservation of foraminifera at the North Atlantic sites is typically good to excellent (fig. S2), and thus is superior to many other deep ocean locations from which $\delta^{18}O$ data have been generated, and we observed no relationship between foraminifer preservation state or burial depth and Δ_{47} temperature (figs. S1 and S2). Our Pleistocene Δ_{47}-based temperatures of 0.0° to 3.4°C (fig. S4) are in agreement with observed present-day temperatures (~3°C) and reconstructed glacial temperatures [−1°C; (*20*)], further attesting to the reliability of the clumped isotope temperatures.

The deviation of Δ_{47}-based temperatures from $\delta^{18}O_b$-based temperatures could imply that deep ocean $\delta^{18}O_{sw}$ was higher than previously

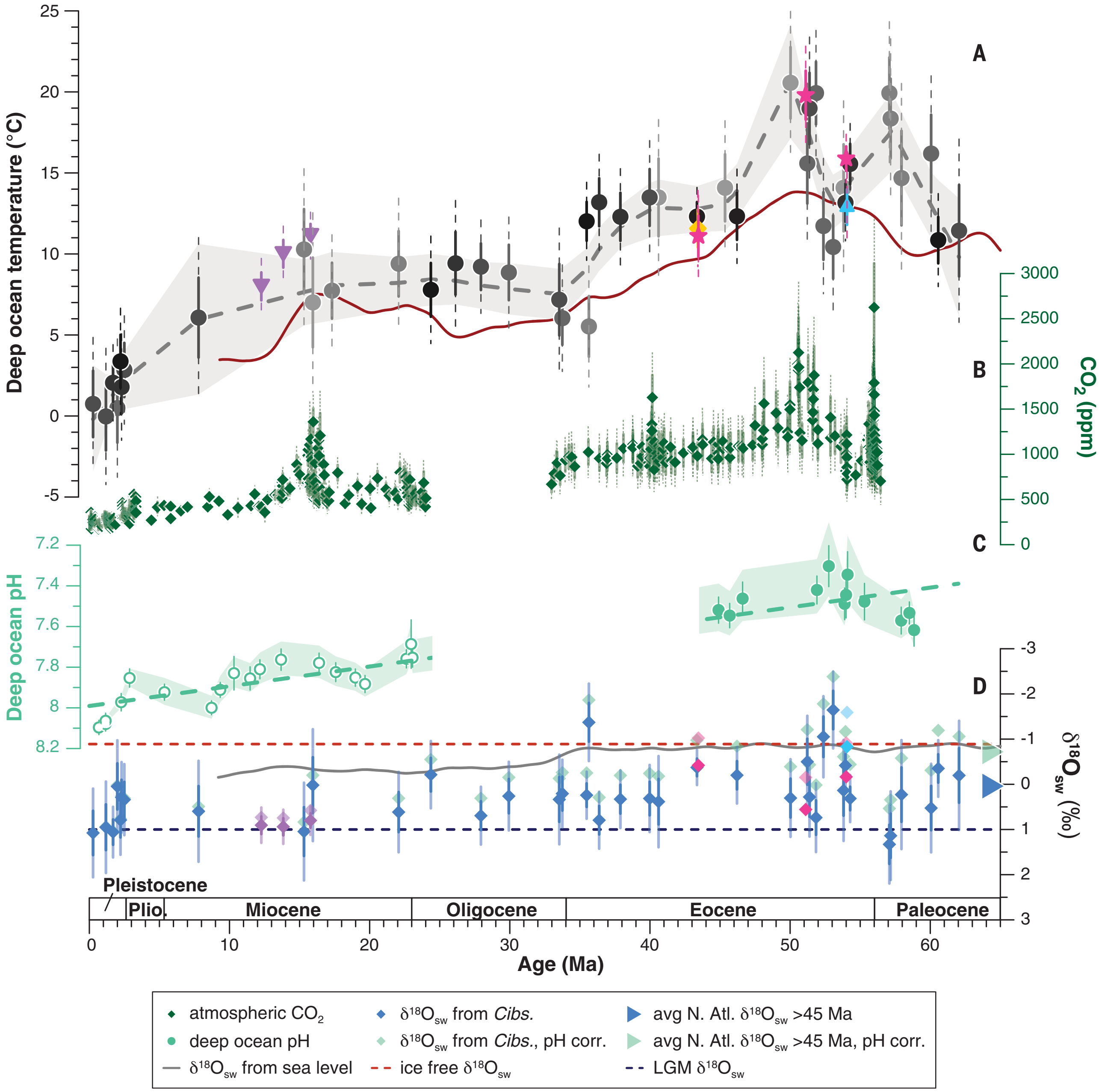


Fig. 2. Cenozoic evolution of seawater isotopic composition and comparison with changes in the carbon cycle. (**A** to **C**) Δ_{47} temperatures as in Fig. 1 (A) with atmospheric CO_2 concentration from boron isotopes in planktic foraminifera (*24*) with 1 and 2 SD (stippled lines) uncertainties (B), and reconstructed deep ocean pH from boron isotopes in benthic foraminifera [(*40*), open symbols; this study, closed symbols] with linear regression (C). Error bars show the influence of uncertainties in $\delta^{11}B$ and temperature at 2 SDs and the shaded area reflects the influence of $\delta^{11}B_{sw}$ ± 0.5‰ (see the materials and methods). (**D**) Calculated $\delta^{18}O_{sw}$ (against the SMOW standard; blue/purple symbols with 68 and 95% confidence intervals) compared with suggested "ice-free" $\delta^{18}O_{sw}$ of −0.9 ‰ [red stippled line; (6)], $\delta^{18}O_{sw}$ of Pleistocene glacial periods of 1‰ (dark blue stippled line), and $\delta^{18}O_{sw}$ reconstructed from sea-level estimates and the estimated ice-free $\delta^{18}O_{sw}$ [gray line; (6)]. Blue data points are from the North Atlantic, and other colors reflect other sites (see Fig. 1). Light green and other faded color symbols are $\delta^{18}O_{sw}$ estimates after correcting $\delta^{18}O_b$ for the long-term increase in deep ocean pH (regression line in C) and assuming a sensitivity of 1.42 ‰ per pH unit (*26*). Although this first-order correction does not eliminate the variability in $\delta^{18}O_{sw}$, it brings the values on |average in closer agreement with the estimated ice-free $\delta^{18}O_{sw}$ for the early part of the record. Arrows on the right show the average North Atlantic $\delta^{18}O_{sw}$ values for >45 Ma before (blue) and after (light green) pH correction.

thought (Fig. 2D, blue symbols versus gray line), for example, because of intensified crustal interaction with seawater due to faster seafloor spreading or from substantial continental storage of ^{16}O in ice sheets or as groundwater. Reconstructions of seafloor spreading rates show large variations but no consistent trend across the Cenozoic (*21*). $\delta^{18}O_{sw}$ values, calculated from Δ_{47} temperatures, regularly approach 1‰ in the North Atlantic and other locations [Fig. 2D (*16*)]. If wholly explained by continental ice sheets, then this would imply that ice sheets reached sizes similar to those in the Last Glacial Maximum, which seems implausible. Nevertheless, uncertainties in the size of Cenozoic ice sheets translate into uncertainties in $\delta^{18}O_{sw}$ and thus could contribute to the discrepancies in temperature estimates from Δ_{47} and $\delta^{18}O_b$. Enhanced storage of ^{16}O-enriched groundwater could also have increased $\delta^{18}O_{sw}$ and has been proposed for the Cretaceous greenhouse [e.g., (*22*)], but the magnitude of this effect is uncertain (*23*). Parts of the ocean interior could at times have been filled with more saline ^{18}O-enriched water. Overall, any of these factors or a combination of them could have contributed to elevating $\delta^{18}O_{sw}$ during much of the Cenozoic, but their potential influence remains unclear.

Instead of an underestimation of $\delta^{18}O_{sw}$, another possibility is that the incorporation of the $\delta^{18}O$ signal in foraminifera could be affected by nonthermal influences such as ocean pH. On the basis of new and published boron isotope data ($\delta^{11}B$) from benthic foraminifera, we reconstructed a pH increase of ~0.6 pH units in the deep ocean over the course of the Cenozoic (Fig. 2C), corresponding to a substantial decrease in atmospheric CO_2 concentration [Fig. 2B (*24, 25*)]. Although evidence for a pH effect on $\delta^{18}O_b$ has not yet been determined experimentally (*7*), the theoretical effect of 1.42‰ per pH unit (*26*), as observed in planktic foraminifera (*27*), would bias early Cenozoic $\delta^{18}O_b$-based temperatures by ~4°C. A pH influence on Δ_{47}, if present, would bias those temperatures in the same direction, but the effect should be approximately four times smaller than for $\delta^{18}O_b$-derived temperatures (*28*, *29*), and has so far not been observed in the typical oceanic pH range (*30*). Although we cannot conclusively demonstrate a pH effect on benthic $\delta^{18}O_b$, it is intriguing that correcting $\delta^{18}O_b$ for the overall Cenozoic trend in deep ocean pH, assuming the same effect as in planktic species (Fig. 2D, light green symbols), brings the calculated $\delta^{18}O_{sw}$ into closer agreement with previous estimates based on sea-level reconstructions (*6*): The average $\delta^{18}O_{sw}$ for >45 Ma becomes −0.72‰ versus 0.04‰ without pH correction (Fig. 2D). This improved agreement points to a pH effect being a likely explanation for the long-term disagreement between $\delta^{18}O_b$- and Δ_{47}-based temperatures. Renewed efforts should focus on investigating pH effects on $\delta^{18}O$ and Δ_{47} in benthic foraminifera to test this hypothesis.

Superimposed on the general trend in Δ_{47} temperatures, some intervals, such as the late Eocene and especially the late Paleocene to early Eocene, stand out with substantial temperature variability that is not apparent to the same extent in our $\delta^{18}O_b$ data or the global $\delta^{18}O_b$ record (Fig. 1 A and B). Although the late Eocene variability currently hinges on one anomalously cold data point at ~35.6 Ma, which could correspond to a previously reported late Eocene cooling event in the North Atlantic and elsewhere (*31*), the temperature variations in the late Paleocene to early Eocene appear to be of longer duration and are better constrained. There, we observe multi-million-year-long temperature swings with a first warming of ~4°C between 58 and 57 Ma, followed by cooling of ~4°C between 57 and 54 Ma, further ~3°C cooling between 54 and 53 Ma, and a second, major warming of 7° to 8°C at 52 Ma (Fig. 1 and fig. S4). Distinct cold spells around 55 Ma have previously been reported for the North Atlantic region (*32*), and the Δ_{47} temperatures from (*18*) show that these extended to the deep South Atlantic (Fig. 1A, pink stars and blue triangle).

The early Eocene is characterized by strong orbital-scale climate variability depicted by variations in $\delta^{18}O_b$ (*2*) of up to 1‰ (4° to 5°C using traditional assumptions), which, in principle, could have been aliased by our lower-resolution Δ_{47} sampling. However, the $\delta^{18}O_b$ data that we obtained with the Δ_{47} measurements do not show anything resembling the scale of variability suggested by Δ_{47}. This discrepancy between Δ_{47} and $\delta^{18}O_b$ is reflected in the substantial temporal variability in calculated $\delta^{18}O_{sw}$. Given the large variations in atmospheric CO_2 of up to 1000 ppm during the early Eocene [Fig. 2B (*25*)], it is likely that deep ocean pH changed more frequently and substantially over this interval than our $\delta^{11}B$ record shows. Such changes in pH could dampen any high-amplitude temperature variance recorded in $\delta^{18}O_b$. However, pH-induced dampening can likely not explain the full scale of the difference between $\delta^{18}O_b$ and Δ_{47}, because such short-lived pH changes would have to have been of similar or larger magnitude as the increase observed over the whole Cenozoic era. Changes in crustal exchange and ice volume are similarly unlikely causes for these comparatively rapid fluctuations in $\delta^{18}O_{sw}$ under peak greenhouse warmth. Instead, the early Paleogene $\delta^{18}O_{sw}$ signal is best explained by changes in deep ocean salinity, with colder temperatures between 54 and 52 Ma accompanied by fresher deep water (low $\delta^{18}O_{sw}$) and peak temperatures at 52 to 49 Ma and 57 to 60 Ma associated with high $\delta^{18}O_{sw}$ values, implying saltier deep water.

Tectonic activity and seafloor spreading in the northern North Atlantic intensified at ~57 Ma, diminished at 52 Ma, and intensified again at 48 Ma (*33*), which may have modified Arctic–North Atlantic oceanic connections, with potentially major effects on Atlantic Ocean circulation. The transitions to colder and fresher deep water sometime between 57 and 54 Ma and back to warm salty deep water at 52 Ma also roughly coincide with a marked increase and subsequent decrease in $\delta^{13}C_b$ (Fig. 1C) that may reflect changes in ocean circulation. Early Eocene changes in ocean circulation are also indicated by nitrogen and sulfur isotope records [fig. S7 (*34, 35*)], suggesting shifts from low-latitude warm, oxygen-poor intermediate, deepwater mass sources to high-latitude, colder, oxygen-rich sources. These latter biogeochemical tracers suggest stepwise changes rather than the transient temperature swings revealed by Δ_{47}, possibly because they reflect intermediate rather than deepwater masses and/or changes in other ocean basins that might differ from those in the deep Atlantic. Regardless, multiple lines of evidence indicate large-scale reorganizations in water mass structure in the Atlantic, and possibly the global ocean, during the early Paleogene.

The observed switches between warm and cold deepwater masses with inferred accompanying changes in salinity imply the need to reassess the interpretation of $\delta^{18}O_b$ records. Temperature and salinity are the key determinants of seawater density; given that both influence $\delta^{18}O_b$, $\delta^{18}O_b$ has previously been invoked as an integrated proxy for seawater density (*36*) rather than for either of the underlying parameters in isolation. Calculating density from $\delta^{18}O_b$ is, however, complicated by large variations in the relationship between $\delta^{18}O_{sw}$ and salinity (*36*). Nonetheless, at any given depth in the deep ocean, density is likely more stable than temperature or salinity alone, which could explain why $\delta^{18}O_b$ does not record the temperature swings that we observed with Δ_{47}. During the early Eocene, latitudinal temperature gradients were reduced [e.g., (*8*)], salinity gradients were enhanced because of a stronger hydrological cycle (*37*), and shallow tropical seas were more widespread. Under these conditions, saline water masses originating at low to mid latitudes could have been more prevalent at times. Density differences may have been small enough to allow for the coexistence of, and switches between, water masses of very different temperatures and/or salinities, as has been proposed for the Paleocene-Eocene Thermal Maximum [e.g., (*38*)]. On time scales shorter than possible variations in the global ocean water isotope budget, the overall isotope budget must stay balanced, calling for freshening in other parts of the ocean during times when the deep Atlantic Ocean was filled with saltier water. Our data suggest that the early Eocene greenhouse was characterized by a more dynamic deep water mass structure than was previously thought.

The strong temperature variations that we observed in the deep Atlantic during the early Eocene are not reflected in the contemporaneous sea surface temperature records available (fig. S5). Although there is a geographic scarcity of records and a potential for large geographic differences in sea surface temperatures (*39*), at face value, this discrepancy between deep ocean and sea surface temperatures lends support to changes in the structure of deep water masses being the main cause of the strong deep ocean temperature variations rather than wholesale swings in global mean temperature. However, the pronounced warming at 52 Ma coincides with a marked increase in reconstructed atmospheric CO_2 [Fig. 2B (*24, 25*)], calling for increased focus on detailed temperature reconstructions from both the surface and deep ocean across this important time interval.

Whereas our data confirm the first-order Cenozoic cooling trend in the deep ocean, they also imply that our canonical views on the relationship of $\delta^{18}O_b$ with ice volume and temperature do not hold true through the Cenozoic, implying that $\delta^{18}O_b$ should not be used as a direct temperature proxy without further constraints. If our warmer deep ocean temperatures turn out to be representative of a large fraction of the global ocean, then this would lead to a higher climate sensitivity to atmospheric CO_2 during Cenozoic greenhouse climates (*5*, *18*, *25*) compared with previous $\delta^{18}O_b$-based estimates. Our finding of major variations in deep ocean temperature during the early Eocene that are not apparent in $\delta^{18}O_b$ points to additional, non-ice volume–related changes in the oxygen isotope composition of seawater potentially in combination with changes in deep ocean pH masking the temperature effect in $\delta^{18}O_b$. Together, these constraints on deep ocean properties emerging from the application of clumped isotope thermometry offer the opportunity to investigate, not only overall greenhouse climate states, but also climate (in)stability under such warmer-than-modern boundary conditions.

REFERENCES AND NOTES

1. J. Zachos, M. Pagani, L. Sloan, E. Thomas, K. Billups, *Science* **292**, 686–693 (2001).
2. T. Westerhold *et al.*, *Science* **369**, 1383–1387 (2020).
3. D. J. Lunt *et al.*, *Clim. Past* **17**, 203–227 (2021).
4. K. D. Burke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13288–13293 (2018).
5. G. N. Inglis *et al.*, *Clim. Past* **16**, 1953–1968 (2020).
6. B. S. Cramer, K. G. Miller, P. J. Barrett, J. D. Wright, *J. Geophys. Res.* **116** (C12), C12023 (2011).
7. T. M. Marchitto *et al.*, *Geochim. Cosmochim. Acta* **130**, 1–11 (2014).
8. C. J. Hollis *et al.*, *Geosci. Model Dev.* **12**, 3149–3206 (2019).
9. M. E. Katz *et al.*, *Paleoceanography* **18**, 1024 (2003).
10. A. Tripati, J. Backman, H. Elderfield, P. Ferretti, *Nature* **436**, 341–346 (2005).
11. C. H. Lear, H. Elderfield, P. A. Wilson, *Science* **287**, 269–272 (2000).
12. C. H. Lear, E. M. Mawbey, Y. Rosenthal, *Paleoceanography* **25**, PA4215 (2010).
13. J. M. Eiler, *Earth Planet. Sci. Lett.* **262**, 309–327 (2007).
14. S. Bernard, D. Daval, P. Ackerer, S. Pont, A. Meibom, *Nat. Commun.* **8**, 1134 (2017).
15. D. Evans *et al.*, *Nat. Commun.* **9**, 2875 (2018).
16. S. E. Modestou, T. J. Leutert, A. Fernandez, C. H. Lear, A. N. Meckler , *Paleoceanogr. Paleoclimatol.* **35**, e2020PA003927 (2020).
17. T. J. Leutert, S. Modestou, S. M. Bernasconi, A. N. Meckler, *Clim. Past* **17**, 2255–2271 (2021).
18. T. Agterhuis, M. Ziegler, N. J. de Winter, L. J. Lourens, *Commun. Earth Environ* **3**, 39 (2022).
19. T. J. Leutert *et al.*, *Geochim. Cosmochim. Acta* **257**, 354–372 (2019).
20. N. Thiagarajan, A. V. Subhas, J. R. Southon, J. M. Eiler, J. F. Adkins, *Nature* **511**, 75–78 (2014).
21. R. D. Müller, M. Sdrolias, C. Gaina, B. Steinberger, C. Heine, *Science* **319**, 1357–1362 (2008).
22. J. E. Wendler, I. Wendler, C. Vogt, J. Kuss, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **441**, 430–437 (2016).
23. A. Davies *et al.*, *Cretac. Res.* **112**, 104445 (2020).
24. J. W. B. Rae *et al.*, *Annu. Rev. Earth Planet. Sci.* **49**, 609–641 (2021).
25. E. Anagnostou *et al.*, *Nat. Commun.* **11**, 4436 (2020).
26. R. E. Zeebe, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 49–57 (2001).
27. H. J. Spero, J. Bijma, D. W. Lea, B. E. Bemis, *Nature* **390**, 497–500 (1997).
28. A. K. Tripati *et al.*, *Geochim. Cosmochim. Acta* **166**, 344–371 (2015).
29. W. F. Guo, *Geochim. Cosmochim. Acta* **268**, 230–257 (2020).
30. J. W. Tang, M. Dietzel, A. Fernandez, A. K. Tripati, B. E. Rosenheim, *Geochim. Cosmochim. Acta* **134**, 120–136 (2014).
31. K. K. Śliwińska, E. Thomsen, S. Schouten, P. L. Schoon, C. Heilmann-Clausen, *Sci. Rep.* **9**, 4458 (2019).
32. M. L. Vickers *et al.*, *Nat. Commun.* **11**, 4713 (2020).
33. C. Gaina, J. Jakob, *Tectonophysics* **760**, 136–151 (2019).
34. V. C. F. Rennie *et al.*, *Nat. Geosci.* **11**, 761–765 (2018).
35. E. R. Kast *et al.*, *Science* **364**, 386–389 (2019).
36. J. Lynch-Stieglitz, W. B. Curry, N. Slowey, *Paleoceanography* **14**, 360–373 (1999).
37. J. Zhu *et al.*, *Earth Planet. Sci. Lett.* **537**, 116164 (2020).
38. A. Tripati, H. Elderfield, *Science* **308**, 1894–1898 (2005).
39. P. M. J. Douglas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6582–6587 (2014).
40. R. Greenop *et al.*, *Clim. Past* **13**, 149–170 (2017).
41. A. N. Meckler *et al.*, Cenozoic clumped isotope temperature record from the deep North Atlantic for: Cenozoic evolution of deep ocean temperature from clumped isotope thermometry, Version 1.0, EarthChem (2022); https://doi.org/10.26022/IEDA/112213.
42. A. N. Meckler *et al.*, Clumped isotope deep-sea temperature data in the South Atlantic for the Early Eocene Climatic Optimum from Meckler *et al.* (2022) for: Cenozoic evolution of deep ocean temperature from clumped isotope thermometry, Version 1.0, EarthChem (2022); https://doi.org/10.26022/IEDA/112215.
43. A. N. Meckler *et al.*, Cenozoic clumped isotope temperature record from the deep North Atlantic for: Cenozoic evolution of deep ocean temperature from clumped isotope thermometry, Pangaea (2022); https://doi.org/10.1594/PANGAEA.945578.

ACKNOWLEDGMENTS

We thank N. Irvali, R. Tapia, L. Griem, J. Donn Holl, L. Al-Saadi, V. Taylor, E. Alagoz, and E. Vinje Galaasen for technical help during sample preparation and clumped isotope analyses in Bergen; A. Fernandez Bremer for sharing Matlab code for error propagation; I. J. Kocken for processing the Utrecht clumped isotope data archive with his R script; A. Roozendaal for sample preparation in Utrecht; A. E. van Dijk for technical support in the Utrecht lab; and M. Kaminski and E. Littley for assistance with boron isotope analyses. This research used samples and data from the International Ocean Discovery Program (IODP) and the Ocean Drilling Program (ODP). The manuscript benefited from the constructive comments of the reviewers. **Funding:** This work was supported by the Swiss National Science Foundation (MHV fellowship to A.N.M.); the European Research Council (starting grant 638467 to A.N.M. and starting grant 805246 to J.W.B.R.); the Trond Mohn Foundation (starting grant BFS2015REK01 to A.N.M.); the Norwegian Research Council (infrastructure grant 245907 to A.N.M.); the Natural Environment Research Council (NERC grant NE/P019331/1 to P.F.S.); the Dutch Research Council (NWO VIDI project 016.161.365 to M.Z.); the Heising Simons Foundation (grant 2022-3314 to A.T.); and the Netherlands Earth System Science Centre (NESSC) (L.J.L.). **Author contributions:** Conceptualization: A.N.M., P.F.S., A.T., M.Z., S.M.B.; Funding acquisition: A.N.M., P.F.S., M.Z., L.J.L., J.W.B.R., A.T., S.M.B.; Initiation: A.N.M., A.T.; Investigation: A.N.M., P.F.S., A.M.P., T.J.L., J.M., T.A., M.Z., L.J.L., J.B., J.W.B.R.; Methodology: A.N.M., M.Z., S.M.B.; Resources: P.F.S., J.W.B.R., L.J.L.; Writing – original draft: A.N.M., P.F.S., M.Z., J.W.B.R.; Writing – review & editing: A.N.M., P.F.S., T.J.L., J.M., M.Z., T.A., L.J.L., J.W.B.R., J.B., A.T., S.M.B. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Replicate-level Δ_{47} data with standard data and detailed data-processing information are available from the EarthChem database (*41*, *42*). Sample-average Δ_{47} temperatures, $\delta^{18}O_b$, and $\delta^{18}O_{sw}$ data are available in the Pangaea database (*43*) and are available in the supplementary materials as data S1. The boron isotope data from benthic foraminifera are available in the supplementary materials as data S3 and will be uploaded to the EarthChem and Pangaea databases.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk0604
Materials and Methods
Figs. S1 to S7
Tables S1 and S2
References (*44*–*79*)
Data S1 to S3

Submitted 18 June 2021; accepted 26 May 2022
10.1126/science.abk0604

NATURAL HAZARDS

Global fast-traveling tsunamis driven by atmospheric Lamb waves on the 2022 Tonga eruption

Tatsuya Kubota[1]*, Tatsuhiko Saito[1], Kiwamu Nishida[2]

On 15 January 2022, the Hunga Tonga-Hunga Ha'apai volcano erupted, producing tsunamis worldwide including first waves which arrived more than 2 hours earlier than what is expected for conventional tsunamis. We investigated the generation and propagation mechanisms of the tsunami "forerunner," and our simulation found that fast-moving atmospheric Lamb waves drove the leading sea height rise whereas the scattering of the leading waves related to bathymetric variations in the Pacific Ocean produced subsequent long-lasting tsunamis. Tsunamis arriving later than the conventionally expected travel time are composed of various waves generated from both moving and static sources, which makes the tsunami, due to this eruption, much more complex and longer-lasting than ordinary earthquake-induced tsunamis.

On 15 January, 2022 (universal time coordinated), a massive volcanic eruption occurred on a small, uninhabited island in the South Pacific, Hunga Tonga-Hunga Ha'apai (black triangle in Fig. 1A) (*1*). This submarine volcano stretches 20 km across and is topped by a submarine caldera 5 km in diameter with minor above-water caldera lips (*2*). Satellite-based observations revealed that the volcanic plume reached as high as 58 km into the mesosphere (*3*) and that the change in the island's topography was so substantial that most of the land disappeared (*2*). Atmospheric pressure increases with amplitudes of ~2 hectopascal (hPa) propagated globally, as observed by the world barograph network (Fig. 1B). Satellite-based observations have captured the global atmospheric wave propagation (*4*, *5*). The initial waves propagating at a velocity of ~300 m/s are Lamb waves, which are atmospheric boundary waves propagating along Earth's surface at the speed of sound (*6–8*) (fig. S1). Similar global Lamb wave propagation has also been recorded during past large volcanic eruptions (*9–11*).

Global coastal tide gauges recorded the tsunamis associated with this eruption (*4*) (fig. S2), but they arrived more than 2 hours earlier than the theoretically predicted tsunami travel time from the volcano, T_{tsun} (green contour lines in Fig. 1A) (*5*). Globally distributed ocean bottom pressure gauges also recorded these fast-traveling tsunamis, with amplitudes of ~3 to 4 hPa at the time corresponding to the Lamb wave arrivals, as well as the subsequent tsunami waves (Fig. 1C). Compared with past major earthquake-induced tsunamis, the ocean bottom pressure changes resulting from this eruption lasted much longer (fig. S3). We investigated the generation mechanism of these enigmatic global forerunning and long-lasting tsunamis related to the atmospheric waves radiated by the eruption.

We studied how a tsunami generated by Lamb waves might propagate across the globe (*8*). Assuming a Lamb wave at a propagation velocity of V_0 = 300 m/s (Fig. 1B), we first conducted a two-dimensional global propagation simulation of atmospheric acoustic gravity waves (Fig. 2A), which include Lamb waves and gravity waves (fig. S1). We then simulated the generation and propagation of ocean waves driven by a moving pressure disturbance (*12*, *13*) using global bathymetry data (*14*). We finally calculated the ocean bottom pressure changes (p_{bot}) as the sum of the atmospheric pressure (p_{atm}) and the sea surface height changes η ($p_{eta} = \rho_w g_0 \eta$, ρ_w, seawater density; g_0, gravity acceleration, 9.8 m/s^2): $p_{bot} = p_{atm} + p_{eta}$ (*13*, *15*).

We compared the observed and simulated waveforms of the barographs and ocean bottom pressure gauges (Fig. 2 and fig. S4). These waveforms coincided well with each other around Lamb wave arrivals. After some additional simulations assuming various V_0 values, we found that the observed tsunamis were reproduced with velocities of V_0 = 280 to 320 m/s (fig. S5).

We show the global propagation of atmospheric pressure changes in movie S1 and Fig. 3, which simulates the Lamb waves and fast-traveling tsunami. In association with the propagation of the ~2 hPa atmospheric pressure increase (Fig. 3A), we can confirm the propagation of the leading sea surface uplift (red arrows in Fig. 3B), which is interpreted as waves forcibly excited by Lamb waves. We also found that ~4 hPa forerunning ocean bottom pressure increased corresponding to the Lamb wave pulse (Fig. 3C). The relative amplitude ratio between the atmospheric and ocean bottom pressures was consistent with global observations (Fig. 1). It should be noted that this wave was not generated by a resonant mechanism between tsunamis and atmospheric waves (*16*) because the sea surface height did not increase with an increase in travel distance. Upon careful inspection of some regions such as the Hawaiian Islands, the South Pacific Islands, and the Mariana Trench, it is noted that scattered tsunamis are generated by these islands and associated steep bathymetry changes, designating them as secondary sources (black arrows in Fig. 3B). A simulation assuming an ocean with a constant water depth—which did not excite these secondary waves—confirmed that the bathymetry-related wave scattering was the cause of the tsunamis following the leading wave (figs. S6 and S7). Wave excitations related to regional-scale bathymetric features such as continental shelves or slopes (*12*, *17*, *18*) should also contribute to these waves.

After the passage of the Lamb-wave pulse and the scattering-originated tsunamis, we can also confirm the propagation of the sea surface depression (blue arrows in Fig. 3B), in which its wavefront spreads at the theoretically predicted tsunami velocity $c_0 = (g_0 h_0)^{0.5}$ (g_0, gravity acceleration; h_0, seawater depth). This wave was expected to be a result of water volume conservation (*15*, *16*) (fig. S8 and movie S2). The sea surface uplift forcibly caused by the Lamb waves propagates at a velocity of V_0, which is accompanied by sea surface subsidence at the source to conserve the total water volume. The subsided sea surface displacement then, as a result of gravity, propagates as tsunamis at velocity c_0. This subsidence wave is one of the factors contributing to tsunamis appearing after the theoretical tsunami arrival, T_{tsun} (component 2a in Fig. 4).

Tsunamis generated around the volcano source region by seafloor crustal deformation resulting from eruptions and caldera collapses along with other mechanisms such as flank failures, sector collapses, and pyroclastic flows (*19–23*) can also contribute to tsunamis after T_{tsun} (component 2c), although these contributions were not considered in our simulation. Another important factor in tsunamis appearing after T_{tsun} is the waves caused by atmospheric waves propagating at a velocity close to that of the tsunamis ($V_0 \sim c_0$, component 2b). Volcanic eruptions excite various atmospheric waves other than Lamb waves (*11*), some of which are referred to as atmospheric gravity wave modes (fig. S1). Some gravity wave modes have a velocity close to that of a tsunami generated by almost any given part of the Pacific Ocean (c_0 ~200 to 220 m/s, h_0 ~4 to 5 km) (*24*, *25*). This results in resonance of the waves (*16*, *24–27*), in which the atmosphere continuously supplies the wave energy to the ocean, causing the tsunami height to increase with the increase in travel distance. All the waves excited by the mechanisms

[1]National Research Institute for Earth Science and Disaster Resilience, Tsukuba, 305-0006, Ibaraki, Japan. [2]Earthquake Research Institute, the University of Tokyo, 113-0032, Tokyo, Japan.
*Corresponding author. Email: kubotatsu@bosai.go.jp

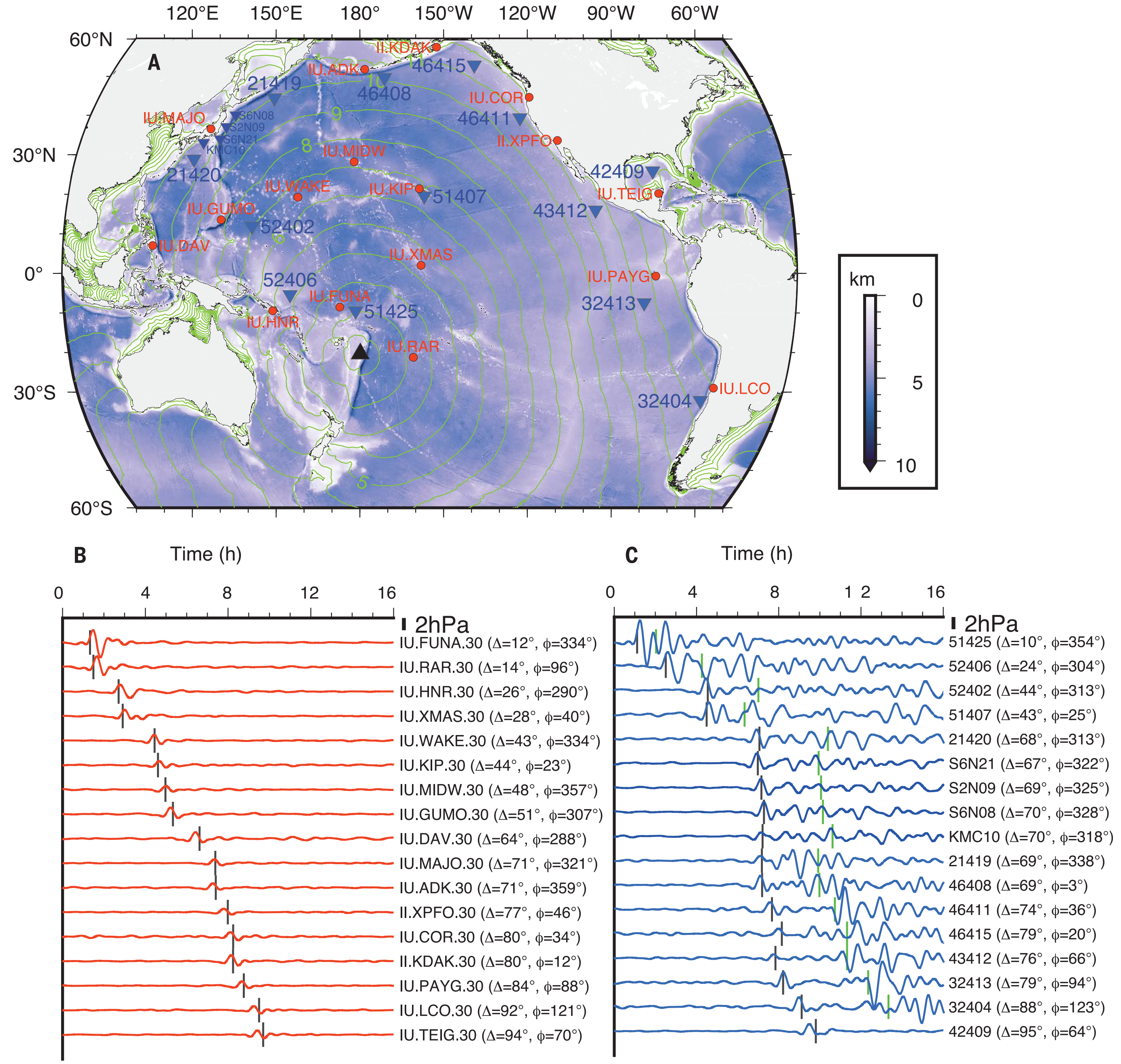


Fig. 1. Location map and observation dataset. (**A**) Location map. The black triangle denotes the Hunga Tonga-Hunga Ha'apai volcano. Red circles denote the barographs. Blue inverted triangles are the ocean bottom pressure gauges. Green contour lines are the theoretical tsunami travel time, T_{tsun}. (**B**) Observed barograms. (**C**) Observed ocean bottom pressure gauge waveforms. The angular distances between the source and stations (Δ) and the azimuth from the source to the station (φ) are also shown. A bandpass filter with a passband of 1800 to 7200s was applied. The theoretical travel times of the Lamb wave (Δ/V_0, black bars) and tsunami (green bars) are also shown.

of components 2a, 2b, and 2c in Fig. 4 propagate at a velocity of c_0 as tsunamis.

To appropriately reproduce the entirety of the observed tsunami waveforms and to quantitatively reveal the entire process of this volcanic eruption and subsequent tsunami generation, it is essential to consider all the contributing factors (components 1, 2a, 2b, and 2c in Fig. 4). Component 2a can be reproduced by the conventional tsunami simulation framework because this component propagates at the tsunami velocity c_0 and the source can be regarded as static. Component 2c can also be essentially reproduced by the conventional simulation, although the details of the tsunami source excited at the volcano have not currently been clarified and more observations and investigations such as field surveys are needed. Even if the amplitudes of the gravity waves are much smaller than those of the Lamb waves, the contribution resulting from gravity waves (component 2b) may be important because of the resonant coupling. To correctly reproduce this component we need to consider various factors such as the source process of the acoustic gravity wave radiation (*10*) and appropriate modeling of the propagation processes including nonlinear effects such as advection related to the jet stream at stratospheric heights (*7*, *28*, *29*). All

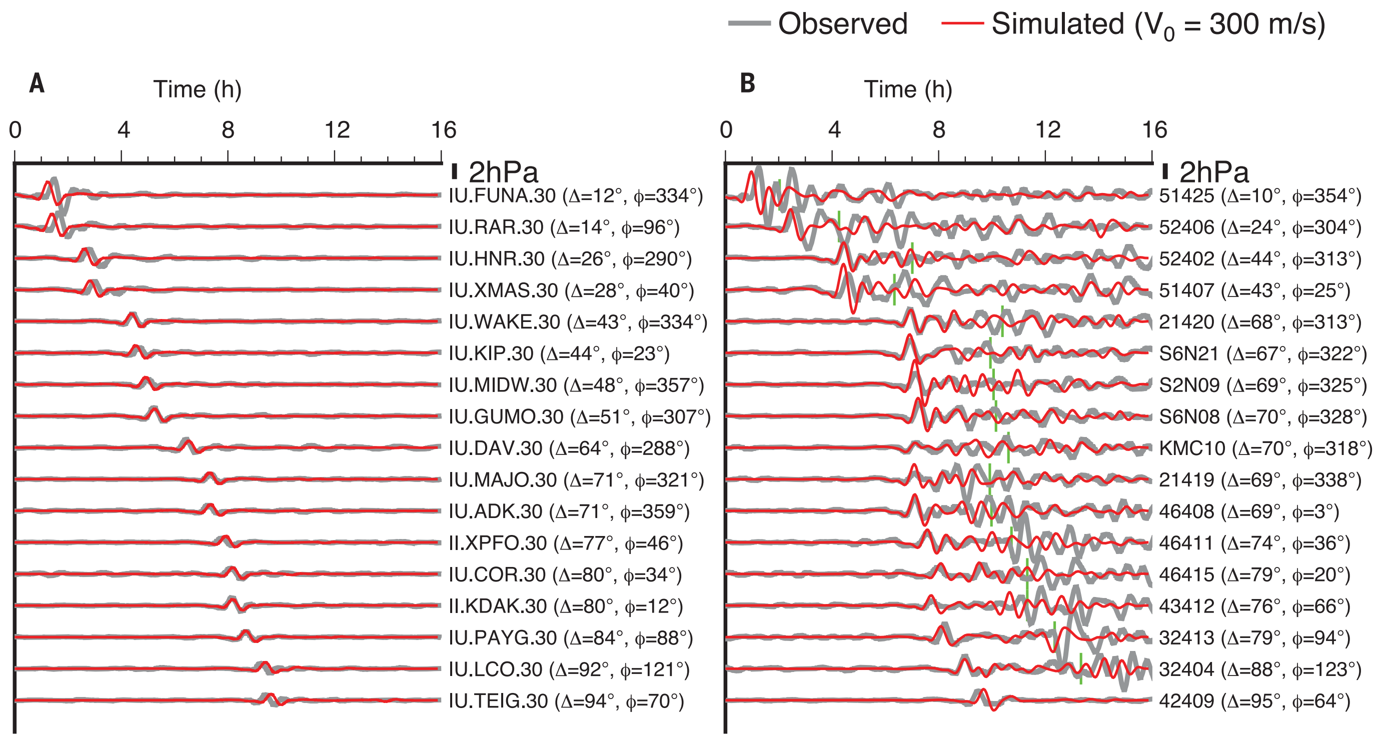


Fig. 2. Comparisons between the observed and simulated waveforms. (**A**) Comparison for the barograph waveforms. (**B**) Comparison for the ocean bottom pressure waveforms. Gray and red traces are the observed and simulated waveforms. The waveform traces are aligned on the basis of the angular distance from the source. The theoretical travel times of the tsunami (green bars) are also shown.

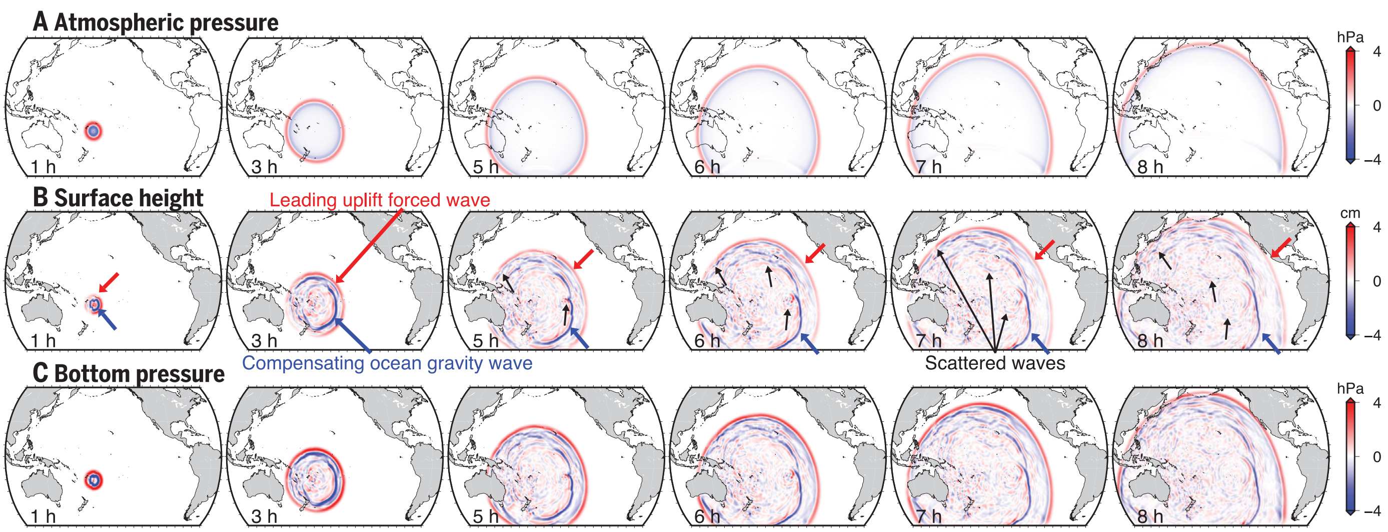


Fig. 3. Snapshots of the generation and propagation of tsunamis. (**A**) Snapshots of the atmospheric pressure change. (**B**) Snapshots of the sea surface height change. Distinct wave signals are marked by arrows. (**C**) Snapshots of the ocean bottom pressure change. An animation of the simulation is available in (*8*) (movie S1).

of these factors should be simulated in future studies.

If the ocean depth is constant, the amplitudes of components 2a and 2c decay with the propagation distance in proportion to $r^{-1/2}$ (r, propagation distance; figs. S9 and S10). By contrast, in component 2b the wave amplitude increases at a rate proportional to $r^{1/2}$ when the resonance occurs between the ocean wave and moving source (figs. S9 and S10), indicating that the contribution of the atmospheric gravity waves (component 2b) can be larger than that of others after propagating a certain distance. These may appear in the tide gauge records at some locations such as the South American coast (fig. S2). The gravity wave modes can contribute to the long durations of the observed waves (fig. S3). However, estimating the contribution of components 2a, 2b, and 2c quantitatively is challenging because these components overlap in the actual records and it is difficult to decompose each contribution.

The characteristics of tsunamis generated by volcanic eruptions, as studied here, render global tsunami warnings more challenging than earthquake-induced tsunamis (*30*, *31*).

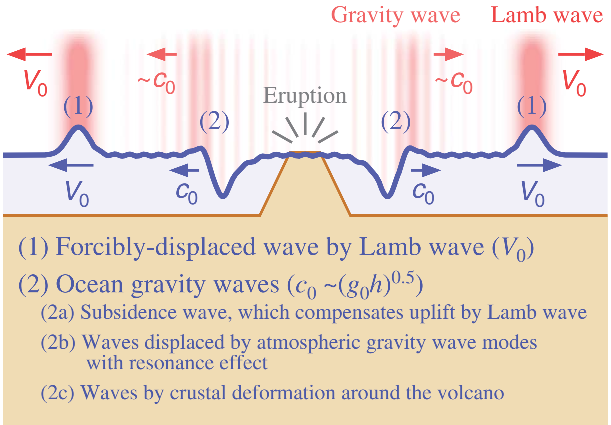


Fig. 4. Schematic illustration of the generation mechanism of the tsunamis in the 2022 Tonga eruption event.

The first challenge is the warning of a tsunami forerunner produced by forced oscillation of Lamb waves. Because the pressure change at the ocean bottom caused by this leading wave is a superposition of the pressure changes caused by the Lamb wave and tsunamis, $p_{bot} = p_{atm} + p_{eta}$ (*13*, *15*), the ocean bottom pressure does not directly represent the tsunami height, whereas the contribution of atmospheric pressure is not contained in the coastal tide gauges. This makes it difficult to accurately calculate the tsunami height using global ocean bottom pressure gauges. We must consider the increase of bottom pressure changes resulting from the effects of the atmospheric pressure to conduct a precise forecast of the global tsunami forerunners. Another critical issue for tsunami warnings is the excitement of tsunamis by atmospheric gravity waves. The atmospheric gravity waves radiated by the eruption amplify the tsunamis with an increase in propagation distance. Even if the amplitudes of the gravity waves are weak, tsunamis are continuously amplified throughout the propagation, which will enlarge the wave amplitude at a greater distance than that of the near-source region. Continuous energy supply from the atmosphere will also increase tsunami duration, up to three days or more, which is much longer than that of ordinary earthquake-induced tsunamis (fig. S3). Several factors control the amplitudes of these global tsunamis, as shown in Fig. 4. However, the most decisive factor in global tsunami size and duration should be how much gravity wave modes with a velocity close to that of tsunamis are excited by the eruption. To precisely forecast long-lasting, resonant-induced tsunamis, we need to accurately model the generation and propagation of gravity wave modes. In the aftermath of disastrous tsunamigenic earthquakes such as those in Sumatra (2004) and Tohoku (2011), we developed global-scale evaluations of tsunami risks and hazards. Our analyses of the 2022 Tonga tsunami shed light on the value of different global tsunami risk evaluations from volcanic eruptions as well as the importance of establishing an understanding of global tsunamis (*32*).

REFERENCES AND NOTES

1. USGS, M 5.8 Volcanic Eruption - 68 km NNW of Nuku'alofa, Tonga (2022); https://earthquake.usgs.gov/earthquakes/eventpage/us7000gc8r/executive.
2. NASA, Dramatic changes at Hunga Tonga-Hunga Ha'apai (2022); https://earthobservatory.nasa.gov/images/149367/dramatic-changes-at-hunga-tonga-hunga-haapai.
3. NASA, Tonga volcano plume reached the mesosphere (2022); https://earthobservatory.nasa.gov/images/149474/tonga-volcano-plume-reached-the-mesosphere.
4. J. Duncombe, *Eos* **103**, (2022).
5. R. G. Andrews, Tonga shock wave created tsunamis in two different oceans. *Sci. Adv.* ada0562 (2022).
6. H. Lamb, *Hydrodynamics* (Dover, ed. 6, 1932).
7. K. Nishida, N. Kobayashi, Y. Fukao, *Geophys. J. Int.* **196**, 312–316 (2014).
8. Materials and methods are available as supplementary materials.
9. R. H. Scott, *Proc. R. Soc. A* **36**, 139–143 (1883).
10. T. Mikumo, B. A. Bolt, *Geophys. J. Int.* **81**, 445–461 (1985).
11. S. Watada, H. Kanamori, *J. Geophys. Res* **115**, B12319 (2010).
12. J. R. Garrett, *Tellus* **22**, 43–52 (1970).
13. T. Kubota, T. Saito, N. Y. Chikasada, O. Sandanbata, *Geophys. Res. Lett.* **48**, e2021GL094255 (2021).
14. National Geophysical Data Center, NOAA, C. Amante, B. W. Eakins, "ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis" (NOAA Tech. Memo. NESDros. Inf. Serv. NGDC-24, 2009); https://www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/docs/ETOPO1.pdf.
15. T. Saito, T. Kubota, N. Y. Chikasada, Y. Tanaka, O. Sandanbata, *J. Geophys. Res. Oceans* **126**, e2020JC017011 (2021).
16. J. Proudman, *Geophys. J. Int.* **2**, 197–209 (1929).
17. R. Vennell, *J. Fluid Mech.* **650**, 427–442 (2010).
18. Y. Tanioka, Y. Yamanaka, T. Nakagaki, *Earth Planets Space* **74**, 61 (2022).
19. B. H. Choi, E. Pelinovsky, K. O. Kim, J. S. Lee, *Nat. Hazards Earth Syst. Sci.* **3**, 321–332 (2003).
20. R. Paris *et al.*, *Nat. Hazards* **70**, 447–470 (2014).
21. S. T. Grilli *et al.*, *Sci. Rep.* **9**, 11946 (2019).
22. R. Williams, P. Rowley, M. C. Garthwaite, *Geology* **47**, 973–976 (2019).
23. R. Paris, *Philos. Trans. R. Soc. A* **373**, 20140380 (2015).
24. F. Press, D. Harkrider, *Science* **154**, 1325–1327 (1966).
25. D. Harkrider, F. Press, *Geophys. J. Int.* **13**, 149–159 (1967).
26. M. Ewing, F. Press, *Trans. Am. Geophys. Union* **36**, 53–60 (1955).
27. P. W. Francis, *J. Volcanol. Geotherm. Res.* **25**, 349–363 (1985).
28. F. P. Bretherton, *Q. J. R. Meteorol. Soc.* **95**, 754–757 (1969).
29. R. S. Lindzen, D. Blake, *J. Geophys. Res.* **77**, 2166–2176 (1972).
30. V. V. Titov *et al.*, *Nat. Hazards* **35**, 35–41 (2005).
31. E. Bernard, V. Titov, Evolution of tsunami warning systems and products. *Philos. Trans. R. Soc. A* **373**, 20140371 (2015).
32. E. L. Geist, H. M. Fritz, A. B. Rabinovich, Y. Tanioka, *Pure Appl. Geophys.* **173**, 3663–3669 (2016).
33. P. Wessel *et al.*, *Geochem. Geophys. Geosyst.* **20**, 5556–5564 (2019).
34. National Research Institute for Earth Science and Disaster Resilience, Ocean bottom pressure data recorded by NIED S-net after Hunga Tonga - Hunga Ha'apai volcano eruption, MOWLAS (2022).
35. National Research Institute for Earth Science and Disaster Resilience, Ocean bottom pressure data recorded by NIED DONET after Hunga Tonga - Hunga Ha'apai volcano eruption, MOWLAS (2022).
36. H. Kubo *et al.*, *Earth Planets Space* **74** (2022).

ACKNOWLEDGMENTS

We thank the contributors of each dataset used in the present study. We used ETOPO1 global bathymetry data for simulation (*14*). Figures were prepared using Generic Mapping Tools (GMT) software (*33*). **Funding:** This work was funded by the following: JSPS KAKENHI JP19K04021 (to T.S.); JSPS KAKENHI JP19K14818 (to T.K.); JSPS KAKENHI JP19H02409 (to T.K. and to T.S.); JSPS KAKENHI JP21K21353 (to T.K., T.S., and K.N.). **Author contributions:** Conceptualization: T.K., T.S., and K.N. Methodology: T.K. and T.S. Investigation: T.K. and K.N. Visualization: T.K. Funding acquisition: T.K. and T.S. Writing – original draft: T.K. Writing – review and editing: T.K., T.S., and K.N. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Global barograph data from the Global Seismic Network (GSN) of the Incorporated Research Institutions for Seismology (IRIS) were downloaded from http://service.iris.edu/irisws/timeseries/1/. Global ocean bottom pressure gauge data from the Deep-ocean Assessment and Reporting of Tsunamis (DART) system of the National Oceanic and Atmospheric Administration (NOAA) are available from https://www.ndbc.noaa.gov/. Ocean bottom pressure gauge data around Japan from the Seafloor observation network for earthquakes and tsunamis along the Japan Trench (S-net) (*34*) and Dense Oceanfloor Network system for Earthquakes and Tsunamis (DONET) (*35*) of National Research Institute of Earth Science and Disaster Resilience (NIED) (*36*) are available from https://doi.org/10.17598/nied.0007-2022-001 and https://doi.org/10.17598/nied.0008-2022-001, respectively. Global coastal tide gauge data were downloaded from the website of the Intergovernmental Oceanographic Commission (IOC) (https://www.ioc-sealevelmonitoring.org).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo4364
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*37*–*39*)
Movies S1 and S2

Submitted 9 February 2022; accepted 2 May 2022
Published online 12 May 2022
10.1126/science.abo4364

NATURAL HAZARDS

Atmospheric waves and global seismoacoustic observations of the January 2022 Hunga eruption, Tonga

Robin S. Matoza[1]*, David Fee[2], Jelle D. Assink[3], Alexandra M. Iezzi[1], David N. Green[4], Keehoon Kim[5], Liam Toney[2], Thomas Lecocq[6], Siddharth Krishnamoorthy[7], Jean-Marie Lalande[8], Kiwamu Nishida[9], Kent L. Gee[10], Matthew M. Haney[11], Hugo D. Ortiz[1], Quentin Brissaud[12], Léo Martire[7], Lucie Rolland[13], Panagiotis Vergados[7], Alexandra Nippress[4], Junghyun Park[14], Shahar Shani-Kadmiel[3], Alex Witsil[2], Stephen Arrowsmith[14], Corentin Caudron[15], Shingo Watada[9], Anna B. Perttu[16,17], Benoit Taisne[16,18], Pierrick Mialle[19], Alexis Le Pichon[20], Julien Vergoz[20], Patrick Hupe[21], Philip S. Blom[22], Roger Waxler[23], Silvio De Angelis[24], Jonathan B. Snively[25], Adam T. Ringler[26], Robert E. Anthony[26], Arthur D. Jolly[27], Geoff Kilgour[28], Gil Averbuch[14], Maurizio Ripepe[29], Mie Ichihara[9], Alejandra Arciniega-Ceballos[30], Elvira Astafyeva[31], Lars Ceranna[21], Sandrine Cevuard[32], Il-Young Che[33], Rodrigo De Negri[1], Carl W. Ebeling[34], Läslo G. Evers[3], Luis E. Franco-Marin[35], Thomas B. Gabrielson[36], Katrin Hafner[37], R. Giles Harrison[38], Attila Komjathy[7], Giorgio Lacanna[29], John Lyons[11], Kenneth A. Macpherson[2], Emanuele Marchetti[29], Kathleen F. McKee[39], Robert J. Mellors[34], Gerardo Mendo-Pérez[40], T. Dylan Mikesell[41], Edhah Munaibari[13], Mayra Oyola-Merced[7], Iseul Park[33], Christoph Pilger[21], Cristina Ramos[42], Mario C. Ruiz[42], Roberto Sabatini[25], Hans F. Schwaiger[11], Dorianne Tailpied[16], Carrick Talmadge[23], Jérôme Vidot[8], Jeremy Webster[22], David C. Wilson[26]

The 15 January 2022 climactic eruption of Hunga volcano, Tonga, produced an explosion in the atmosphere of a size that has not been documented in the modern geophysical record. The event generated a broad range of atmospheric waves observed globally by various ground-based and spaceborne instrumentation networks. Most prominent was the surface-guided Lamb wave (≲0.01 hertz), which we observed propagating for four (plus three antipodal) passages around Earth over 6 days. As measured by the Lamb wave amplitudes, the climactic Hunga explosion was comparable in size to that of the 1883 Krakatau eruption. The Hunga eruption produced remarkable globally detected infrasound (0.01 to 20 hertz), long-range (~10,000 kilometers) audible sound, and ionospheric perturbations. Seismometers worldwide recorded pure seismic and air-to-ground coupled waves. Air-to-sea coupling likely contributed to fast-arriving tsunamis. Here, we highlight exceptional observations of the atmospheric waves.

The 15 January 2022 eruption of Hunga volcano (*1*), Tonga, was an unusually energetic explosive event. This climactic eruption (the largest eruption of an episode) began just after ~04:00 UTC (~17:00 local time) from a submerged vent and delivered volcanic tephra and gas primarily into the stratosphere. An umbrella cloud developed at ~30 km above sea level, with a much higher central transient overshoot. Hunga is a largely submerged massif located ~65 km to the north-northwest of Tongatapu, Kingdom of Tonga. Eruption episodes consisting of relatively low-energy Surtseyan activity in 2009 and 2014–2015 had built a tephra cone that connected the established islands of Hunga Tonga and Hunga Ha'apai on the northwestern portion of the massif (*2*). Surtseyan eruptions transitioned into violent, impulsive eruptions starting on 19 December 2021 as part of the most recent episode. The climactic 15 January eruption produced a broad range of atmospheric waves observed globally by numerous ground-based and spaceborne instrumentation systems, including atmospheric pressure sensors, seismometers, hydrophones, Global Navigation Satellite System (GNSS) receivers, and weather satellites (Fig. 1A) (*3*). Here, we highlight exceptional multi-technology observations of this extraordinary event in the modern digital record and provide initial interpretations of the atmospheric wave types generated and their propagation around the globe.

The onset of the most recent eruptive episode was characterized remotely by seismicity and co-eruptive infrasound on 19 December 2021, preceded by seismic activity on 18 December 2021 (16:49:46 UTC; body-wave magnitude: 4.0) (Fig. 1B) (*3*). Eruptive activity continued until 4 January 2022, with decreasing infrasonic amplitudes at International Monitoring System (IMS) infrasound station IS22 (1848 km range) and intermittent detections by IMS hydroacoustic stations. Powerful eruptive infrasound activity resumed on 13 January 2022, with amplitudes ~10 times that of the December activity. Infrasound continued on 14 January, accompanied by seismic tremor (*3*) (fig. S2, A and B); infrasound amplitudes subsequently decreased, while the number of hydroacoustic *T*-phase detections increased. After brief relative quiescence, at least four IMS hydroacoustic (fig. S3), all 53 IMS infrasound, and numerous seismic stations detected the main climactic eruption on 15 January 2022 [04:14:45 UTC; moment magnitude (M_w): 5.7 to 5.8; table S1]. Regional infrasound, barometer, and volcanic plume observations suggest a complex eruption sequence occurring between 04:00 and ~04:30, not just a single onset or explosion (Figs. 1A, 2E, and 3A). A final major eruption at ~08:31 UTC on 15 January was detected by at least 20 IMS infrasound and two IMS hydroacoustic stations, after which the volcanic activity decreased.

Atmospheric waves (*4*) are propagating mechanical perturbations in the atmospheric

[1]Department of Earth Science and Earth Research Institute, University of California, Santa Barbara, CA, USA. [2]Wilson Alaska Technical Center and Alaska Volcano Observatory, Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA. [3]R&D Department of Seismology and Acoustics, Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands. [4]AWE Blacknest, Brimpton, Reading, UK. [5]Lawrence Livermore National Laboratory, Livermore, CA, USA. [6]Seismology-Gravimetry, Royal Observatory of Belgium, Brussels, Belgium. [7]NASA Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA, USA. [8]CNRM, Université de Toulouse, Météo-France, CNRS, Lannion, France. [9]Earthquake Research Institute, University of Tokyo, Tokyo, Japan. [10]Department of Physics and Astronomy, Brigham Young University, Provo, UT, USA. [11]US Geological Survey (USGS), Alaska Volcano Observatory, Anchorage, AK, USA. [12]Norwegian Seismic Array (NORSAR), Kjeller, Norway. [13]Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, IRD, Géoazur, Sophia Antipolis, Valbonne, France. [14]Roy M. Huffington Department of Earth Sciences, Southern Methodist University, Dallas, TX, USA. [15]Laboratoire G-Time, Department of Geosciences, Environment and Society, Université libre de Bruxelles, Brussels, Belgium. [16]Earth Observatory of Singapore, Nanyang Technological University, Singapore. [17]Volcanic Risk Solutions, Massey University, Palmerston North, New Zealand. [18]Asian School of the Environment, Nanyang Technological University, Singapore. [19]Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), Vienna, Austria. [20]CEA, DAM, DIF, Arpajon, France. [21]BGR (Federal Institute for Geosciences and Natural Resources), Hannover, Germany. [22]Geophysics Group, Los Alamos National Laboratory, Los Alamos, NM, USA. [23]National Center for Physical Acoustics, University of Mississippi, University, MS, USA. [24]School of Environmental Sciences, University of Liverpool, Liverpool, UK. [25]Department of Physical Sciences and Center for Space and Atmospheric Research (CSAR), Embry-Riddle Aeronautical University, Daytona Beach, FL, USA. [26]USGS Albuquerque Seismological Laboratory, Albuquerque, NM, USA. [27]USGS Hawaiian Volcano Observatory, Hilo, HI, USA. [28]GNS Science, Wairakei Research Centre, Taupō, New Zealand. [29]Department of Earth Sciences, University of Florence, Florence, Italy. [30]Departamento de Vulcanología, Instituto de Geofísica, Universidad Nacional Autónoma de México, Mexico City, Mexico. [31]Université de Paris, Institut de Physique du Globe de Paris, Paris, France. [32]Vanuatu Meteorology and Geohazards Department, Port Vila, Vanuatu. [33]KIGAM (Korea Institute of Geoscience and Mineral Resources), Daejeon, Korea. [34]Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA. [35]Volcanological Observatory of the Southern Andes (OVDAS), National Geology and Mining Service (SERNAGEOMIN), Temuco, Chile. [36]Graduate Program in Acoustics, Penn State University, State College, PA, USA. [37]Incorporated Research Institutions for Seismology (IRIS), Washington, DC, USA. [38]Department of Meteorology, University of Reading, Reading, UK. [39]NASA Goddard Space Flight Center, Greenbelt, MD, USA. [40]Posgrado en Ciencias de la Tierra, Universidad Nacional Autónoma de México, Mexico City, Mexico. [41]Norwegian Geotechnical Institute (NGI), Oslo, Norway. [42]Instituto Geofísico, Escuela Politécnica Nacional, Quito, Ecuador.
*Corresponding author. Email: rmatoza@ucsb.edu

fluid. Nonlinearities in the propagation cause the spectrum to evolve (i.e., energy cascading) and may result in shock-wave formation and period lengthening. Gravity waves are disturbances to the balance between buoyancy and gravity [frequency (f) $\lesssim$ 3 mHz]; acoustic waves manifest as propagating compressions and rarefactions ($f \gtrsim$ 4 mHz). These different physical mechanisms lead to different propagation speeds. Acoustic-gravity waves (AGWs) are waves exhibiting both buoyant and compressional motion (*5*), typically with millihertz frequencies and long wavelengths (tens of kilometers) relative to density stratification scale heights (fig. S4). Lamb waves (*6*) are AGWs propagating along Earth's surface, with group velocities near the mean sound speed of the lower atmosphere [~310 m/s for a 16-km scale height above Earth's surface (*7*)]. Lamb waves are associated with the largest atmospheric explosions from volcanic eruptions (*8*) and nuclear tests (*9*) and have periods on the order of several to hundreds of minutes. Audible sound refers to higher-frequency acoustic waves that can be heard by humans. Infrasound (*10*) refers to acoustic waves below the standard audio range. The crossover between audible and infrasound is often given as 20 Hz.

Of the atmospheric waves produced by the climactic Hunga explosion, the most prominent is the Lamb wave ($f \lesssim$ 0.01 Hz), which propagates efficiently and is detected globally by numerous ground-based and spaceborne geophysical instrumentation systems (Fig. 1A, fig. S5, and movies S1 to S6). Despite the Lamb wave's large amplitude, its waveform pressure increase as a function of time (rise time) is relatively slow and does not have characteristics of a shock wave. Over 6 days, we observed global propagation of at least four minor-arc Lamb wave passages (A1, A3, A5, and A7) and three (A2, A4, and A6) major-arc (antipodal) passages (Figs. 1A, inset, and 2, A and B, and fig. S6A).

The number of Lamb wave passages observed for Hunga (four, plus three antipodal) is approximately the same as observed for the 1883 Krakatau eruption (*11*, *12*) (Fig. 2A). The exceptional spatiotemporal resolution of the evolving wavefield from 2022 Hunga, in comparison with 1883 Krakatau, is a consequence of more than a century of advances in instrumentation technology and global sensor density (Fig. 1A). Measurements of Lamb wave peak-to-peak pressure amplitudes as a function of distance indicate that the atmospheric pressure pulse generated by the Hunga event is comparable to that of the 1883 Krakatau eruption (*12*) (Fig. 2F and fig. S8). However, the duration of the Krakatau Lamb pulse was ~30% longer than that of Hunga at comparable stations (Fig. 2A). Peak-to-peak pressure amplitudes from Hunga generally decreased logarithmically from 1473 Pa (756 km) with increasing distance (range) from the source (Fig. 2F and fig. S9). We infer that the notable scatter in amplitudes at distances >7500 km is related to winds and wavefront focusing around the spherical Earth (*3*) and to a potentially anisotropic source. The Hunga signal amplitudes are more than an order of magnitude larger than those generated by the 1980 Mount St. Helens eruption (*13*).

Equivalent explosive yields for large volcanic eruptions have previously been estimated using pressure recordings, but quantitative comparisons with nonvolcanic sources are problematic. During the atmospheric nuclear testing era of the 1950s and 1960s, theoretical and empirical relationships were generated relating AGW amplitudes and periods to explosive yield (*14*, *15*). We find that such relationships are inapplicable to the signals generated by Hunga, as they result in unphysically large equivalent yields (*3*) (fig. S10A). This difference is presumably because, for a given energy release, the long-duration climactic eruption excites longer-period pressure disturbances than the near-instantaneous nuclear reaction (fig. S10B). Hunga signals have peak-to-peak pressures comparable to those produced by the largest historical atmospheric nuclear test (58 megatons, USSR, 1961) (*16*), but the dominant eruption signal periods (1700 to 2500 s) are approximately four times longer than those of the anthropogenic explosion (400 to 700 s) (*17*).

The Hunga eruption pressure waves have complex waveform and spectral characteristics, likely related to both source and propagation.

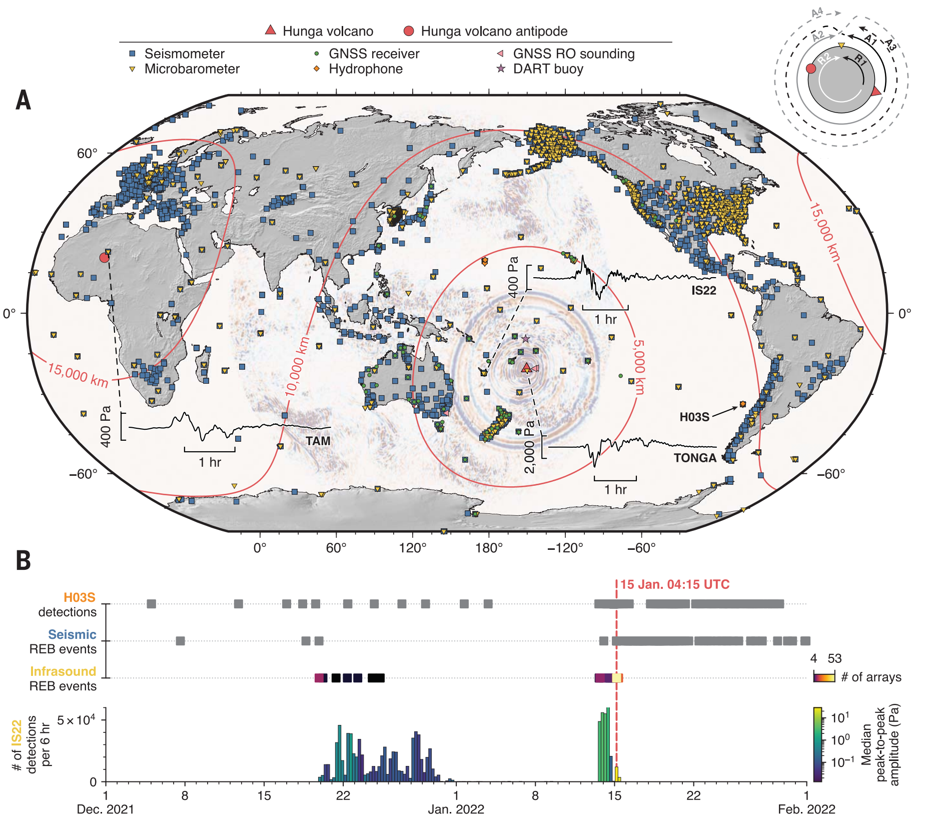


Fig. 1. Global distribution of recording geophysical sensors used in this study and remotely observed eruption chronology. (**A**) Sensor map. Background image is brightness temperature difference (Himawari-8) at 07:10 UTC on 15 January 2022. Selected 4-hour pressure waveforms are filtered from 10,000 to 100 s. Upper-right inset shows Hunga wave paths around Earth. (**B**) Hunga activity, December 2021 through January 2022, observed at IMS hydrophone, seismic, and infrasound stations (REB, Reviewed Event Bulletin); Hunga detections from nearest IMS infrasound array IS22 (1848 km). Frequency responses for atmospheric pressure sensors used in this study are displayed in fig. S1.

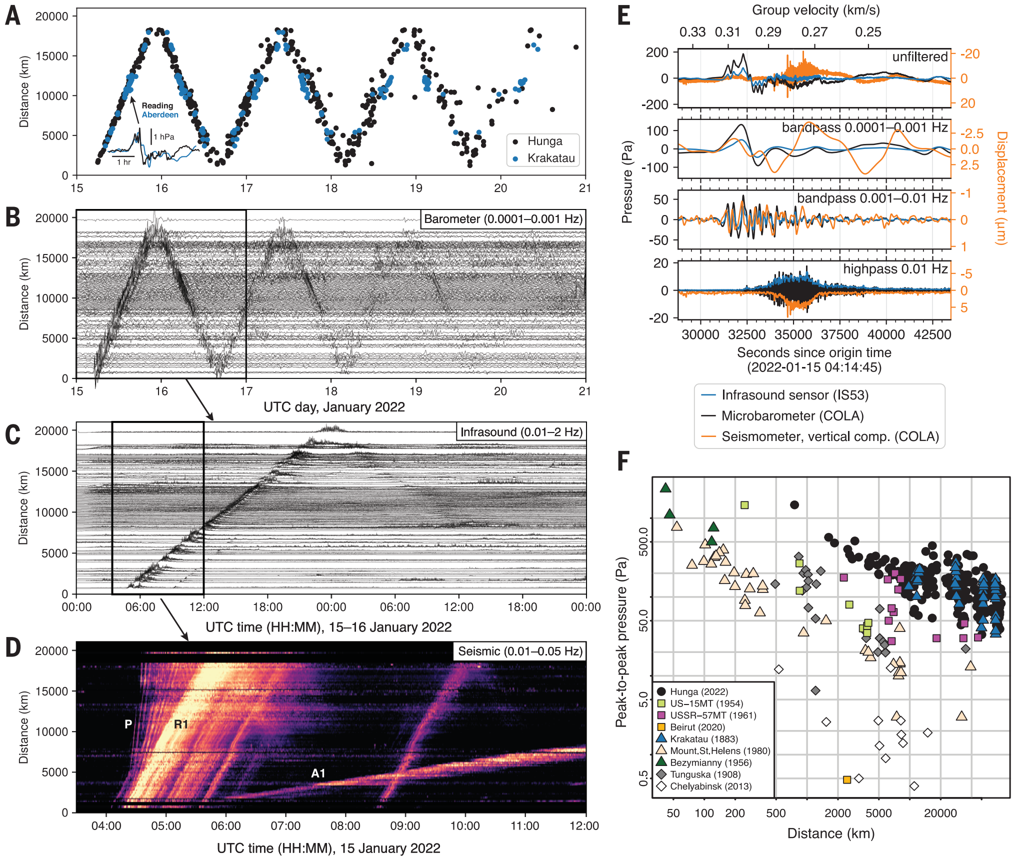


Fig. 2. Ground-based observations. (**A**) Lamb wave arrival times for 2022 Hunga eruption (black) compared with 1883 Krakatau eruption (blue). (Inset) Lamb A1 arrival waveform comparison (*3*). Global record sections of (**B**) barometer, (**C**) infrasound, and (**D**) seismic data showing the multiple arrivals and wave passages (see Fig. 1A, inset); waveforms aggregated by radial distance (fig. S7). A separate Rayleigh R1 is associated with the later ~08:31 event. (**E**) Colocated microbarometer (black), infrasound sensor (blue), and seismometer (orange) waveforms; lower panel shows inverted displacement envelope. (**F**) Wideband peak-to-peak pressure versus distance comparing 2022 Hunga with large historical explosive events (table S2).

The Lamb wave is the largest-amplitude pressure wave arrival (Fig. 2B) (*3*). Near Hunga, the Lamb wave consists of at least two pulses and begins with a 7- to 10-min pressure increase, followed by a second larger compression and subsequent long rarefaction phase (Figs. 1A and 2). This sequence is different from a single bipolar pulse typical of large anthropogenic explosions (*18*). The shallow-submarine volcanic source presumably contributes to this waveform complexity (*19*). The Lamb wave period ranges from 0.3 to 10 mHz (3300 to 100 s), and the group velocity is ~315 m/s (*3*, *20*) (fig. S11). Each subsequent antipodal passage produces an observed 90° phase shift in the Lamb wave (*21*) (fig. S12). This 90° phase shift is expected given the comparison of the asymptotic forms of the equation for a traveling wave on the surface of a sphere from before the antipodal crossing to that from after crossing (*21*, *22*). The Lamb wave is composed of several AGW modes, and the Hunga signals show distinct dispersion at higher frequencies (fig. S13), which was similarly noted for other large AGW signals (*20*). Some barometer observations also show the arrival of a lower-velocity gravity wave (figs. S11 and S14).

The climactic Hunga eruption also produced remarkable long-range infrasound (f ~ 0.01 to 20 Hz), clearly detected at most IMS infrasound arrays (fig. S15) and at numerous regional arrays and networks (*3*) (table S4 and figs. S16 to S21). Infrasound signals arrive after the Lamb wave; at most stations, the Lamb wave dominates below ~0.01 Hz, followed by broadband infrasound (Fig. 3). The IMS infrasound network recorded at least two direct and two antipodal infrasonic wave arrivals from the main explosive event. At most of the infrasound stations, array processing indicates direct infrasonic arrivals for ~2 hours, with group velocities between 250 and 290 m/s (*3*) (fig. S15). Infrasound amplitudes after the first Lamb wave arrival A1 are on the order of several pascals and are observed to decrease with each global wave passage (Fig. 2F). Complex waveform interference effects are observed for stations near the source and the antipode, where the wavetrains of successive arrivals overlap (*3*). Prominent time evolution in signal backazimuth and apparent velocity is observed at many infrasound arrays, especially at stations for which the propagation path crosses the circumpolar vortex (*3*) (fig. S22).

Accounts of audible sound (f > 20 Hz) were reported across Alaska as far as 10,000 km from Hunga [compared with ~4800 km for the 1883 Krakatau eruption (*12*)] and are verified by ~30-min-duration signals on higher-sample-rate low-frequency microphone stations (Fig. 3E). The audio signals arrive after the Lamb wave and at the end of the infrasound wavetrain and consist of short-duration impulsive signals consistent with repeated "booms" reported by observers. Linear propagation and attenuation models cannot explain the high-frequency infrasound and audible sound at these extreme ranges, implying nonlinearity in generating the higher frequencies along the propagation path (*3*, *13*). Evidence of nonlinearity in Fig. 3E is twofold. First, the high-frequency spectral slope during the "peak" time window approximates that of an ideal shock wave in its old-age (*3*) (but still nonlinear) decay: f^{-2}, followed by a faster exponential roll-off at frequencies where atmospheric absorption dominates nonlinearity. Second, the impulsive events, when separated from the lower-frequency, higher-amplitude infrasound portion by filtering (from 10 to 40 Hz), have coarsely sampled N-wave shapes reminiscent of explosions or sonic booms. Substantial increases in global population and advances in societal connectivity (e.g., internet versus telegraph) presumably contribute to the enhanced reports of audibility at distances greater than those historically documented for Krakatau and other large events.

Owing to its extraordinary amplitude, the Lamb wave produced coupled signals at multi-technology stations (Fig. 2E) (*3*). For example,

in the Mediterranean, the Lamb wave produced signals on hydrophones at ~50 m water depth near Stromboli volcano, 17,740 km from Hunga (*3*) (fig. S17B).

Seismometers worldwide recorded ground motions associated with both pure seismic waves (figs. S2 and S23) and air-to-ground coupled atmospheric waves (Fig. 3 and figs. S24 and S6B). We associate the most prominent seismic (*P*, *S*, and Rayleigh waves) and atmospheric arrivals (Fig. 2) with the main eruption at 04:14:45 UTC, which had a reported M_w of between 5.7 and 5.8. Our observations of multiple overlapping seismic phases (Fig. 2D) suggest a longer-duration source process, with at least two discrete events and multiple phases. Additionally, seismic ground motions globally exhibit a marked spectral peak at 3.7 mHz (Fig. 3D). We interpret this peak as Rayleigh-wave propagation (corresponding to Earth normal mode $_0S_{29}$) resulting from the coupling of fundamental acoustic mode oscillations of the atmosphere near the volcanic source into the solid Earth (*3*) (fig. S25). This solid Earth mode was also excited during the 1991 eruption of Mount Pinatubo (*23*); however, the seismic oscillations generated by the climactic 2022 Hunga eruption are more than an order of magnitude larger (*3*).

Numerous additional Earth observation systems recorded the atmospheric waves from the climactic eruption. Data from neutral atmospheric radio occultations (ROs), satellite-based radiometers, and dual-frequency GNSS receivers, in conjunction with data from ground-based infrasound stations and a DART (Deep-ocean Assessment and Reporting of Tsunamis) buoy (1225 km), reveal strong seismo- and hydroatmospheric coupling in the aftermath of the eruption (Fig. 4). The Lamb wave arrival time at IS-II (station CTAO, 3997 km) is consistent with that obtained using brightness temperature differences measured by the Himawari-8 satellite (*3*) (fig. S5). At this time, an RO profile over Eastern Australia (RO-III, 3781 km) clearly displays heightened gravity wave activity in the stratosphere. In the hours after the eruption, ROs in the vicinity of Hunga (RO-I, 366 km, and RO-II, 453 km) also reveal strong gravity wave activity in the stratosphere with temperature perturbations of ±4 K, four times the typical background activity.

The atmospheric waves also propagated to the ionosphere, where 1 Hz data recorded in real time by ground-based GNSS stations can be converted to ionospheric total electron content (TEC). TEC data clearly demonstrate wave-like structures of unprecedented magnitude traveling between ~320 and 1000 m/s. TEC profiles (G-I and G-II) collocated with infrasound stations IS-I (station MSVF, 756 km) and IS-II show the arrival of the Lamb wave in the ionosphere ~24 min after it is recorded at the infrasound station (propagating at an apparent vertical velocity of ~312 m/s for an assumed ionospheric shell height of 450 km). As in the global barometer data (Fig. 2B), the Lamb wave was observed worldwide in TEC data. In addition, a DART buoy (B-I) and a nearby TEC record (G-III) north of Hunga record tsunami-like waves generated by the atmospheric pulse [i.e., air-sea waves (*8*)], 1 hour before the appearance of tsunami signatures of direct volcanic origin (*3*) (fig. S26).

Understanding these geophysical observations from the Hunga eruption requires accurate propagation modeling. However, simulating atmospheric wave propagation is challenging here for multiple reasons. (i) The complexity of the highly energetic, shallow-submarine, and multiphase eruption is beyond existing capability for modeling the source and the subsequent repartition of energy among the different waves (*3*). (ii) The physical problem involves multiple scales. Indeed, observed atmospheric waves contain energy extending from the acoustic-gravity regime, including a strong Lamb wave, through the infrasonic range, and into audio frequencies (Fig. 3). (iii) Atmospheric wave propagation is strongly nonlinear, which leads to energy cascading into higher frequencies even far from the event. For such energetic events, wave propagation nonlinearities remain important far from the source. Considering (ii) and (iii) together, the challenge is due to the nonlinear energy cascading that couples these various regimes (acoustic-gravity, infrasound, audio)

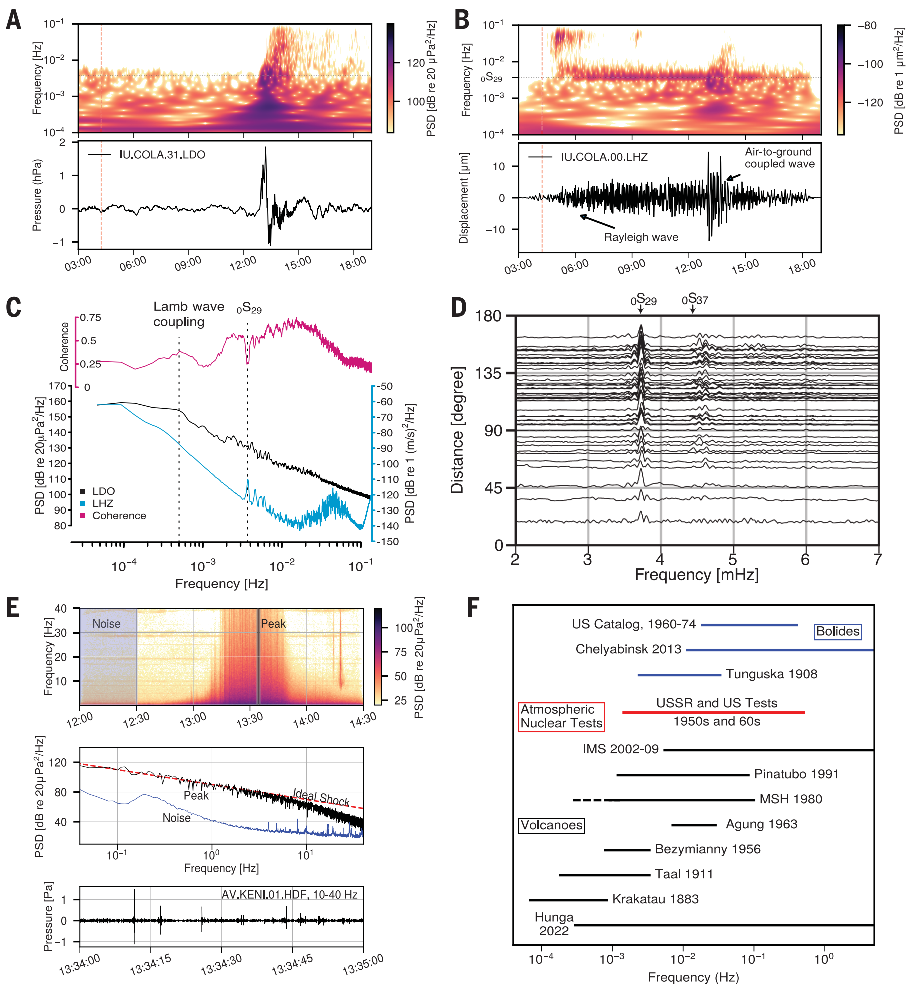


Fig. 3. Seismoacoustic spectral properties. Colocated wideband (**A**) pressure and (**B**) seismic spectrograms (top) and unfiltered waveforms (bottom). (**C** and **D**) Power spectral densities (PSD) and seismoacoustic coherences worldwide show that pressure waves couple to the solid Earth through both (i) direct conversion as the Lamb wave passes the station and (ii) near-source excitation of atmospheric acoustic modes. (**E**) Alaska infrasound stations recorded audio range signals at great distances, apparent in the spectra (top) and as intermittent transients with shock-like features (middle and bottom panels). (**F**) Observed wideband pressure spectral character of the Hunga eruption compared with published instrumental observations of previous events (table S3).

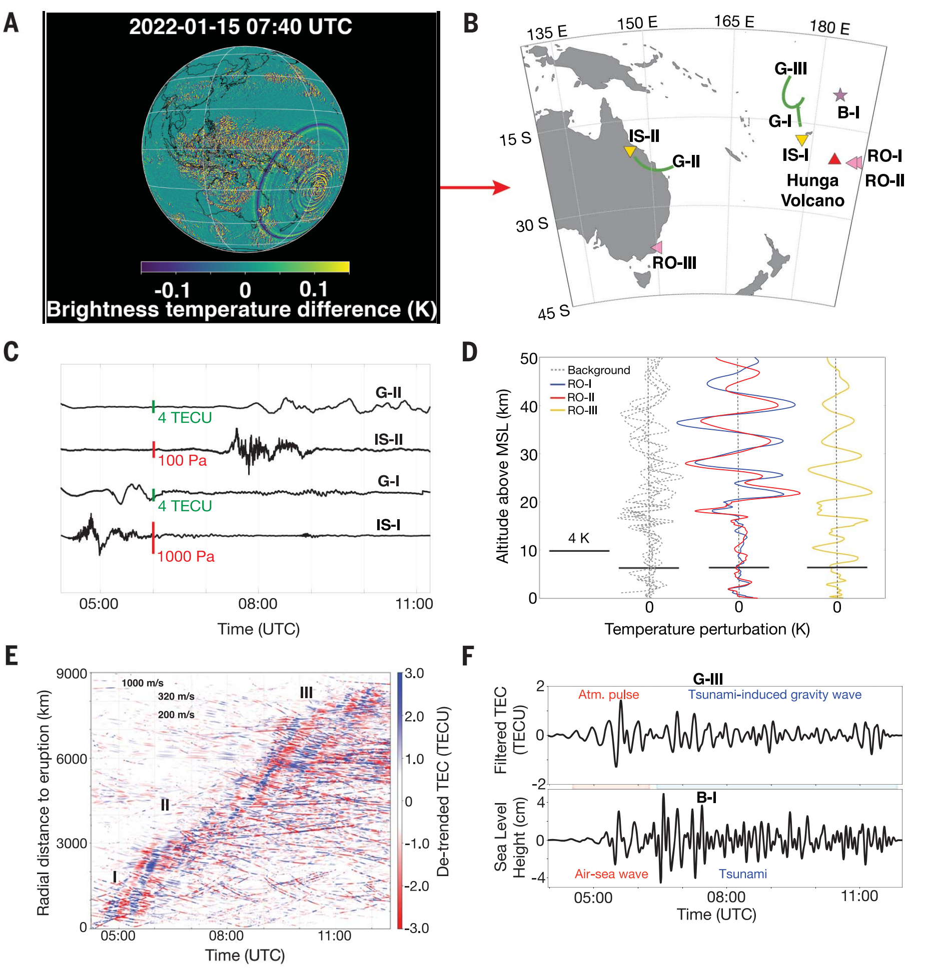


Fig. 4. Seismo- and hydroatmospheric coupling from Earth's surface to space. (**A**) Brightness temperature variations in Himawari-8 data showing waves emanating from the Hunga eruption site. (**B**) Map of the inset in (A), with measurement locations in this figure. Ionospheric pierce point arcs (see supplementary materials section 1.13) are shown in green for the Lamb wave arrival for links G-I and G-II, and from 04:00 to 12:00 UTC for link G-III. (**C**) Infrasound (stations IS-I and IS-II) and TEC (GNSS links G-I and G-II) waveforms showing Lamb wave arrival; all signals high-pass filtered with 0.278 mHz (corresponding to 1-hour period) cutoff. (**D**) RO-I and RO-II at 06:50 UTC and 10:00 UTC showing strong coherent gravity wave activity several hours after the eruption; RO-III at 07:42 UTC also exhibits large gravity waves coincident with Himawari-8 data (A). (**E**) Hodochron plot of TEC records showing long-distance ionospheric wave propagation after the eruption. Features I and II are the first arrivals with different apparent wave velocities (551 to 1333 m/s) due to the near-field wavefront curvature. Feature III, identified more than 6000 km from the eruption, propagates at 478 m/s and is more likely linked to long-period gravity waves. (**F**) Buoy B-I data compared with TEC data from an adjacent GNSS link (G-III) showing efficient air-sea-air coupling across a broad frequency spectrum (*3*).

and requires modeling methods that account for that coupling. (iv) Finally, substantial temporal and spatial variations of atmospheric conditions along propagation paths render a stratified atmospheric model inappropriate. Existing propagation algorithms [based, for instance, on the equations of fluid mechanics, the parabolic approximation of the wave equation, normal-mode summation, or ray tracing (*3*)] are limited in their physics and computational feasibility (fig. S27). Nevertheless, preliminary simulations (*3*) find notable departures of predicted propagation paths from great circle paths (fig. S28 and movie S7), which leads to direction-of-arrival deviations qualitatively in agreement with observations (fig. S22 and table S5).

The impacts of volcanic atmospheric waves are usually limited, but sometimes shock waves from strong volcanic explosions damage nearby infrastructure (*24*, *25*). Atmospheric waves from the main Hunga eruption had far more extensive effects. Unusual sea level changes or tsunamis were observed in the Pacific earlier than predicted, as well as in the Caribbean and Mediterranean without direct ocean routes. We report observations of early sea level oscillations in the Pacific (*3*). At coastal tide gauges, the tsunami onset time approximately coincides with the Lamb wave arrival (2 hPa pulse); the tsunami onset is unclear, but wave amplitudes gradually increase over 2 to 4 hours to >1 m in some locations. In contrast, deep-sea tsunamimeters record a clear leading 5 hPa pressure pulse, more than double that of the air-pressure pulse (*3*) (figs. S29 and S30). Air-sea coupling (*8*, *26*) likely caused these exceptional observations and should be considered in future scenarios for tsunami early-warning systems.

Geophysical records of the January 2022 Hunga eruption represent an unparalleled global dataset of atmospheric wave generation and propagation, providing an opportunity for multi-technology observation, modeling, and validation that is unprecedented in the modern record. The datasets highlighted here are not exhaustive; there is outstanding potential for augmenting details of the global wavefield capture by incorporating numerous additional interdisciplinary datasets, including citizen-science data (*27*, *28*). The January 2022 Hunga eruption presents an extraordinary opportunity to advance understanding of rarely captured physical phenomena, including global Lamb wave propagation, atmospheric free-oscillations coupling with the solid Earth, nonlinear energy cascading in atmospheric wave propagation, excitation of infrasound and audible sound at global distances, air-sea waves, and many others.

REFERENCES AND NOTES

1. The name of the volcano is "Hunga," not "Hunga Tonga." "Hunga" refers to the entire volcano, rather than the islands. "Hunga Tonga" refers specifically to the most southwestern of the two islands.
2. R. G. Vaughan, P. W. Webley, *J. Volcanol. Geotherm. Res.* **198**, 177–186 (2010).
3. Materials and methods and data and network citations are available as supplementary materials.
4. E. E. Gossard, W. H. Hooke, *Waves in the Atmosphere: Atmospheric Infrasound and Gravity Waves—Their Generation and Propagation* (Elsevier, 1975).
5. K. C. Yeh, C. H. Liu, *Rev. Geophys.* **12**, 193–216 (1974).
6. H. Lamb, *Proc. R. Soc. London Ser. A* **84**, 551–572 (1911).
7. F. P. Bretherton, *Q. J. R. Meteorol. Soc.* **95**, 754–757 (1969).
8. D. Harkrider, F. Press, *Geophys. J. Int.* **13**, 149–159 (1967).
9. A. D. Pierce, J. W. Posey, *Geophys. J. R. Astron. Soc.* **26**, 341–368 (1971).
10. R. Matoza, D. Fee, D. Green, P. Mialle, in *Infrasound Monitoring for Atmospheric Studies: Challenges in Middle Atmosphere Dynamics and Societal Benefits*, A. Le Pichon, E. Blanc, A. Hauchecorne, Eds. (Springer, Cham, 2019), pp. 1023–1077.
11. R. H. Scott, *Proc. R. Soc. London* **36**, 139–143 (1883).

12. R. H. Strachey, in *The Eruption of Krakatoa and Subsequent Phenomena*, Report of the Krakatoa Committee of the Royal Society, G. J. Symons, Ed. (Trübner and Co., 1888).
13. J. W. Reed, *J. Geophys. Res.* **92**, 11979–11992 (1987).
14. J. W. Posey, A. D. Pierce, *Nature* **232**, 253–253 (1971).
15. G. B. Olmsted, "Detection of airborne low-frequency sound from nuclear explosions," United States Air Force, Office for Atomic Energy, Operation Castle Report WT-931 (1955)
16. W. L. Donn, D. M. Shaw, A. C. Hubbard, *IEEE Trans. Nucl. Sci.* **10**, 285–296 (1963).
17. A. D. Pierce, C. A. Moo, J. W. Posey, "Generation and propagation of infrasonic waves," Air Force Cambridge Research Laboratories, AFCRL-TR-73-0135 (1973).
18. D. Fee *et al*., *J. Geophys. Res. Atmos.* **118**, 6122–6143 (2013).
19. J. J. Lyons, M. M. Haney, D. Fee, W. G. Wech, C. F. Waythomas, *Nat. Geosci.* **12**, 952–958 (2019).
20. F. Press, D. Harkrider, *J. Geophys. Res.* **67**, 3889–3908 (1962).
21. J. N. Brune, J. E. Nafe, L. E. Alsop, *Bull. Seismol. Soc. Am.* **51**, 247–257 (1961).
22. J. H. Ansell, *Geophys. J. Int.* **32**, 95–117 (1973).
23. S. Watada, H. Kanamori, *J. Geophys. Res.* **115**, B12319 (2010).
24. K. Kato, H. Yamasato, *Earth Planets Space* **65**, 2 (2013).
25. G. A. Valentine, *J. Volcanol. Geotherm. Res.* **87**, 117–140 (1998).
26. J. Proudman, *Geophys. Suppl. Mon. Not. R. Astron. Soc.* **2**, 197–209 (1929).
27. A. M. Portas, L. Barnard, C. Scott, R. G. Harrison, *Philos. Trans. R. Soc. London Ser. A* **374**, 20150223 (2016).
28. E. Calais *et al*., *Science* **376**, 283–287 (2022).

ACKNOWLEDGMENTS

We thank M. Garces, A. Wech, and two anonymous reviewers for their comments that improved the manuscript. L.M., A.K., and S.K. thank A. Moore at JPL for her advice on processing GNSS data. The views expressed herein are those of the authors and do not necessarily reflect the views of the CTBTO Preparatory Commission. The views expressed in the article do not necessarily represent the views of the US Department of Energy or the US government. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US government. This product article has been peer reviewed and approved for publication consistent with USGS Fundamental Science Practices (https://pubs.usgs.gov/circ/1367/). Distribution statement: Cleared for release. **Funding:** Funding was provided by National Science Foundation grant EAR-1847736 (R.S.M., H.D.O., and R.D.N.); Defense Threat Reduction Agency Nuclear Arms Control Technology program under contract HQ003421F0112 (D.F., L.T., A.W., and K.A.M.); US Geological Survey Alaska Volcano Observatory (D.F., M.M.H., J.L., and H.F.S.); US Geological Survey Hawaiian Volcano Observatory (A.D.J.); National Science Foundation grant EAR-1901614 (D.F.); National Science Foundation grant EAR-1952392 (A.M.I.); US Department of Energy by the LLNL under contract DE-AC52-07NA27344 (K.K.); Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA (S.K., L.M., P.V., A.K., and M.O.-M.); French Agence Nationale de la Recherche (ANR) under reference ANR-19-CE04-0003 and CNES for APR project UVTECGEOX (L.R. and E.Mu.) and CNES APR project "RealDetect" (E.A.); JSPS KAKENHI Grant 19K04034 (S.W.); JSPS KAKENHI Grant Number 21K21353 (K.N., S.W., and M.I.); Royal Society Grant IES\R2\202007 (S.D.A.); Los Alamos National Laboratory award number 89233218CNA000001 (P.S.B. and J.W.); Projects PAPIIT-UNAM-NI108219 and UCMEXUS-CONACYT (A.A.-C.); Strategic Science Investment Funding to GNS Science, New Zealand, within the Hazards and Risk Management Programme (G.K.); internal funding from Penn State College of Engineering (T.B.G.); NASA Earth Surface and Interior Program under Ben Phillips, under a grant to R. Kahn at the NASA Goddard Space Flight Center (K.F.M.); UCMEXUS-CONACYT grant 699758 (G.M.-P.); Proyecto Generación de capacidades para la difusión de alertas tempranas y para el desarrollo de instrumentos de decisión ante las amenazas sísmicas y volcánicas dirigidos al Sistema Nacional de Gestión de Riesgos (M.C.R. and C.R.); Basic Research Project of KIGAM GP2020-017 and GP2021-006 (I.-Y.C. and I.P.); SAGE II–IDA Network Operations, SU-19-1001-08-UCSD and IGPP/SIO/UCSD internal funding (C.W.E. and R.J.M.); DARPA Cooperative Agreement HR00112120003 (J.B.S. and R.S.). This research was partly supported by the Earth Observatory of Singapore (EOS) through its funding from the National Research Foundation of Singapore and the Singapore Ministry of Education under the Research Centres of Excellence initiative. This work constitutes EOS contribution number 432 (B.T., D.T., and A.B.P.). This research was funded by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development (NNSA DNN R&D). **Author contributions:** R.S.M. designed and led the research. A.M.I. coordinated the supplementary materials. R.S.M., D.F., J.D.A., A.M.I., D.N.G., K.K., L.T., T.L., S.K., J.-M.L., K.N., K.L.G., M.M.H., H.D.O., Q.B., L.M., L.R., P.V., A.N., J.P., S.S.-K., A.W., S.A., C.C., S.W., A.B.P., B.T., P.M., A.L.P., J.Ve., P.H., P.S.B., R.W., S.D.A., J.B.S., A.T.R., R.E.A., A.D.J., G.K., G.A., M.R., M.I., and E.Mu. drafted the manuscript and processed and visualized the data as displayed in the main manuscript; L.T., P.S.B., and J.-M.L. provided the supplementary movies. All authors took part in discussions and data analysis, contributed to the manuscript, contributed to the supplementary materials, performed a full interactive review of the original manuscript, and approved the submitted version. Authors Matoza (R.S.M.) through Ichihara (M.I.) are listed in approximate order of contribution; all other authors [Arciniega-Ceballos (A.A.-C.) through Wilson (D.C.W.)] are listed alphabetically. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** A complete list of the data and materials availability and data acknowledgements can be found in section 1.15 of the supplementary materials (*3*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo7063
Materials and Methods
Supplementary Text
Figs. S1 to S30
Tables S1 to S5
References (*29*–*88*)
Movies S1 to S7
Data S1 to S8

Submitted 26 February 2022; accepted 26 April 2022
Published online 12 May 2022
10.1126/science.abo7063

NANOMATERIALS

Locking volatile organic molecules by subnanometer inorganic nanowire-based organogels

Simin Zhang[1], Wenxiong Shi[2], Xun Wang[1]*

The intermolecular forces among volatile organic molecules are usually weaker than water, making them more difficult to absorb. We prepared alkaline earth cations–bridged polyoxometalate nanoclusters subnanometer nanowires through a facile room-temperature reaction. The nanowires can form three-dimensional networks, trapping more than 10 kinds of volatile organic liquids effectively with the mass fraction of nanowires as low as 0.53%. A series of freestanding, elastic, and stable organogels were obtained. We prepared gels that encapsulate organic liquids at the kilogram scale. Through removing solvents in gels by means of distillation and centrifugation, the nanowires can be recycled more than 10 times. This method could be applied to the effective trapping and recovery of organic liquids.

Water-based hydrogels can be easily prepared through the formation of hydrogen bond networks between water molecules and gelators such as polymer chains and/or inorganic nanostructures (*1–3*). Compared with hydrogen bonds, the intermolecular forces among volatile organic molecules are usually weaker, hindering the semisolidification of volatile organic liquids and development of functional organogel materials.

Subnanometer nanowires (*4–7*) can form three-dimensional (3D) networks and gels similar to polymers. We prepared alkaline earth metal cations–bridged polyoxometalate nanoclusters (AE-POM) subnanometer nanowires through a facile room-temperature reaction [supplementary materials (SM), materials and methods 1.2].

First, we prepared calcium-POM (Ca-POM) nanowires. We investigated morphologies of Ca-POM nanowires through transmission electron microscopy (TEM), atomic-resolution aberration-corrected TEM (AC-TEM), atomic force microscopy (AFM), and small-angle x-ray diffraction (SXRD). The nanowires dispersed in octane with a length of several micrometers and were curved and entangled with each other (Fig. 1, A and B, and fig. S1). The diameter could be distinguished from AFM data, which was ~1 nm (Fig. 1, C and D), so the aspect ratio was as high as several thousands. The spacing distance between two nanowires was calculated from the SXRD data as ~3.78 nm (fig. S2). As shown in the atomic-resolution AC high-angle annular-dark field

[1]Laboratory of Organic Optoelectronics and Molecular Engineering, Department of Chemistry, Tsinghua University, Beijing 100084, China. [2]Institute for New Energy Materials and Low Carbon Technologies, School of Materials Science and Engineering, Tianjin University of Technology, Tianjin 300387, China.
*Corresponding author. Email: wangxun@mail.tsinghua.edu.cn

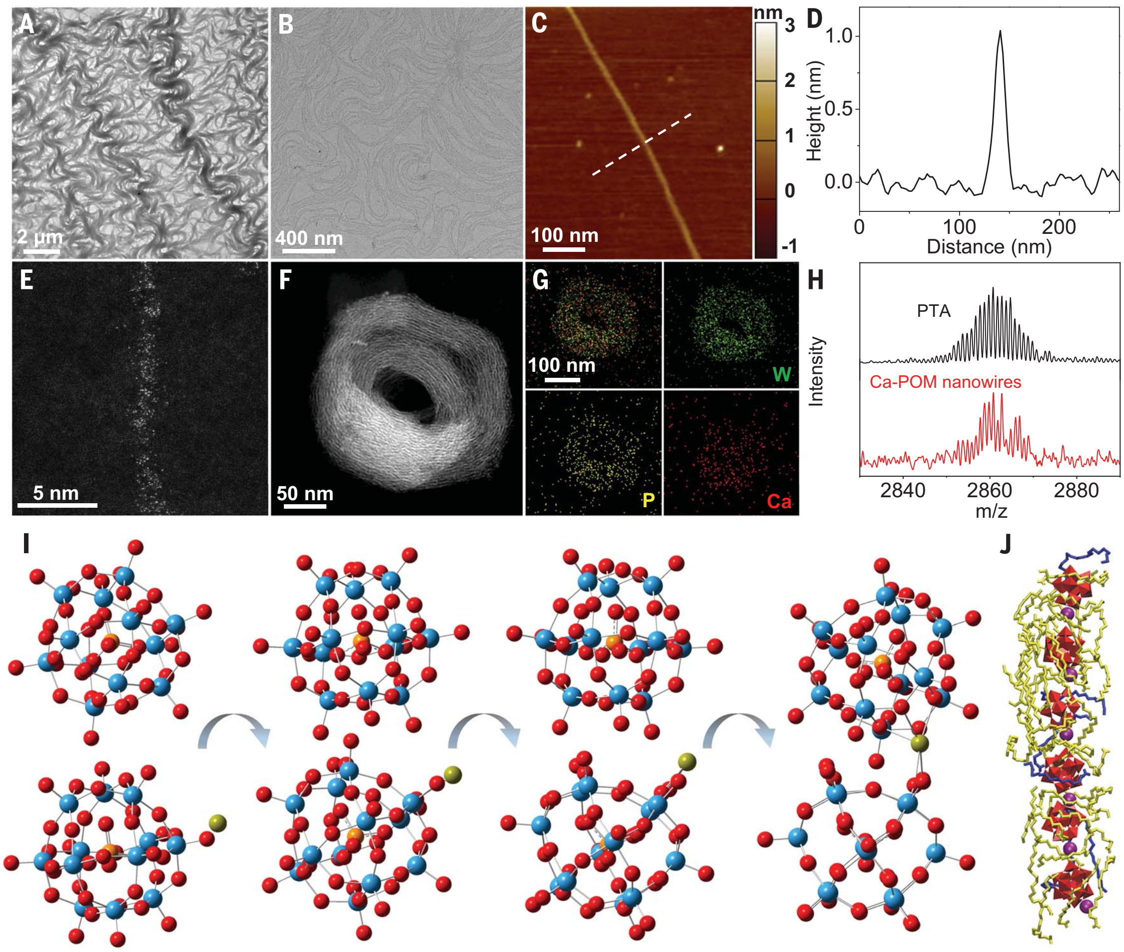


Fig. 1. Morphologies of Ca-POM nanowires. (**A** and **B**) TEM images of nanowires. (**C**) Typical AFM topographic image of the nanowire and (**D**) the line profile along the dashed line. (**E**) AC-HAADF-STEM image of the nanowire. (**F**) STEM image and (**G**) corresponding EDS elemental mapping images of coiled nanowires. (**H**) MALDI-TOF-MS results of PTA and nanowires. (**I**) Geometry optimization of nanowires structure. The red, blue, orange, and yellow balls indicate O, W, P, and Ca, respectively. (**J**) The structural diagram of the nanowire. Purple balls indicate Ca^{2+}, red models indicate $PW_{12}O_{40}{}^{3-}$ nanoclusters, and yellow and blue chains represent oleylamine and protonated oleylamine, respectively.

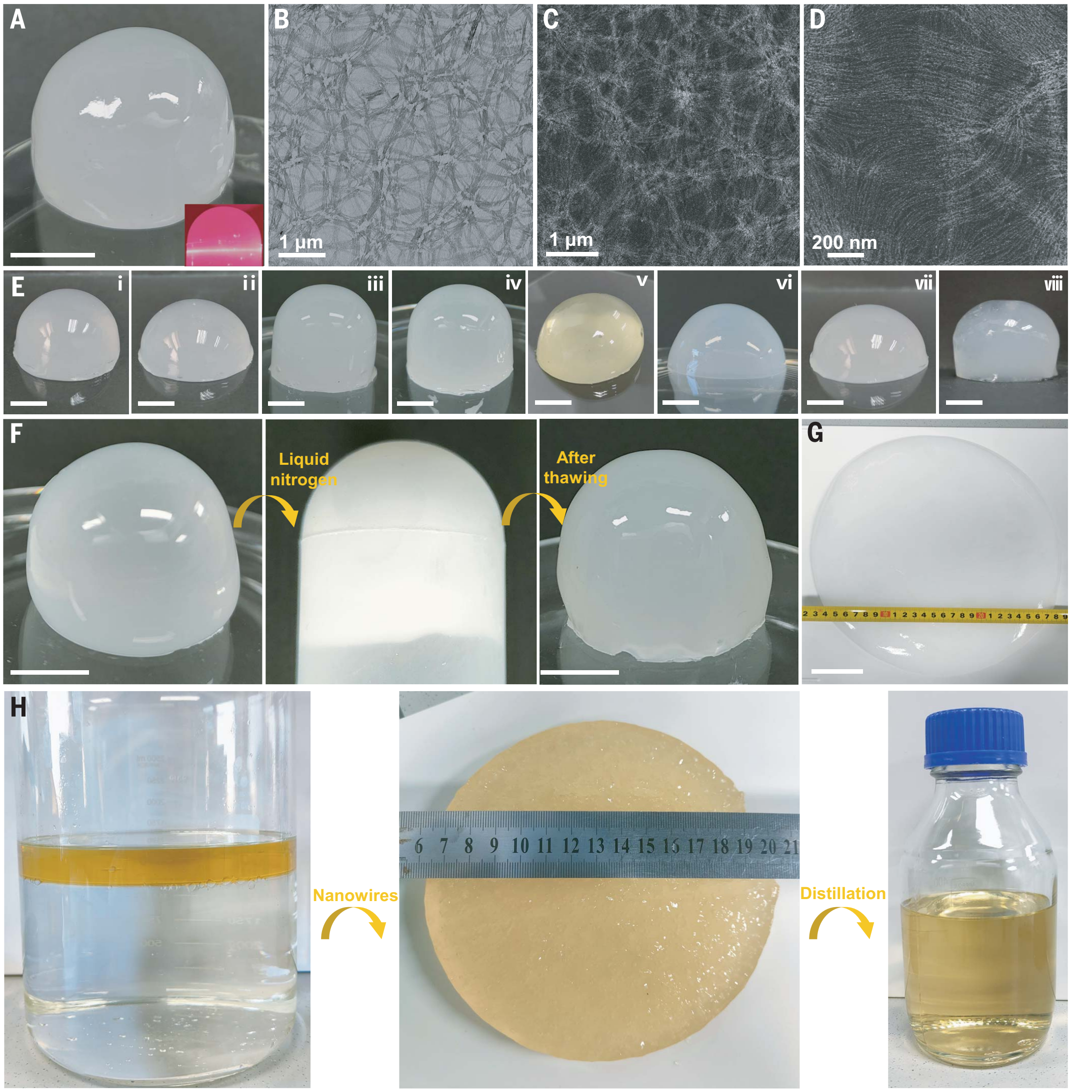


Fig. 2. Ca-POM nanowire–organic liquids gels. (**A**) Photograph of the nanowires-octane gel. (Inset) The Tyndall effect of the gel. (**B**) TEM image and (**C** and **D**) STEM images of nanowire networks. (**E**) Photographs of nanowire-organic liquids gels: (i) cyclohexane, (ii) hexane, (iii) toluene, (iv) 2,5-dimethyl-2,5-di-(tert-butylperoxy) hexane, (v) petrol, (vi) n-propyl ether, (vii) 1-dodecanethiol, and (viii) octadecene. (**F**) Photographs of a nanowire-octane gel frozen in liquid nitrogen and subsequently thawed. The gel was sealed in a tube and then placed in liquid nitrogen. (**G**) A nanowire-octane gel with a diameter of ~26 cm and thickness of ~1.5 cm. (**H**) Photographs of oil spill recovery process with nanowires. The nanowire-petrol gel has a diameter of ~15 cm and thickness of ~2 cm. Scale bars, (A), (E), and (F) 1 cm; (G) 5 cm.

scanning TEM (AC-HAADF-STEM) image (Fig. 1E), the nanowire was mainly composed of nanoclusters. When ethanol was added into the dispersion of nanowires in octane, the nanowires coiled, and some nanorings formed (Fig. 1F and fig. S3), demonstrating their high flexibility. The composition of nanowires was investigated through x-ray photoelectron spectroscopy (XPS) and energy dispersive x-ray spectroscopy (EDS) elemental mapping. Ca-POM nanowires were mainly composed of Ca, tungsten (W), phosphorus (P), and oxygen (O) (fig. S4). The EDS mapping demonstrated that Ca, W, and P dispersed in nanowires uniformly (Fig. 1G). The structure of nanowires was further investigated through Fourier transform infrared (FTIR) spectroscopy and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS). As shown in fig. S5, the FTIR characteristic absorption peaks of phosphotungstic acid (PTA) also appeared in the FTIR spectrum of nanowires (*8*), suggesting that $PW_{12}O_{40}^{3-}$ nanoclusters existed in nanowires. The absorption peaks at 2850 and 2915 cm^{-1} belonged to oleylamine. We further used MALDI-TOF-MS to test PTA and nanowires, revealing the existence of $PW_{12}O_{40}^{3-}$ in nanowires. Because PTA and Ca-POM nanowires were broken into $PW_{12}O_{39}^{-}$ fragments during the ionization process, the peak positions shifted to about 2860 (Fig. 1H and fig. S6) (*9*). The pK_a (–log10K_a, where K_a is the acid dissociation constant) of protonated oleylamine is 10.7, and the pK_a3 of PTA is 2.0 to 3.3 (*10*, *11*). During the experiment, ~0.347 mmol PTA and 9.7 mmol oleylamine were added, so PTA was completely dissociated as H^+ and $PW_{12}O_{40}^{3-}$, and oleylamine was protonated, which was also demonstrated by the FTIR spectra (fig. S7). According to the inductively coupled plasma-atomic emission spectrometry (ICP-AES) data (table S1), the ratio of $PW_{12}O_{40}^{3-}$ to Ca^{2+} in nanowires was about 1:1, so there should be protonated oleylamine with equal molar quantity of $PW_{12}O_{40}^{3-}$ on nanowires to maintain electrical neutrality. Molecular dynamics (MD) simulations (movie S1) and geometry optimization (Fig. 1I and fig. S8) demonstrated that the Ca^{2+} bridged $PW_{12}O_{40}^{3-}$ in the final stable state of nanowires. According to the above characterization results, Ca^{2+} bridged $PW_{12}O_{40}^{3-}$ nanoclusters through electrostatic interaction to form nanowires, and oleylamine and protonated oleylamine attached to nanowires through coordination and electrostatic interactions (Fig. 1J). MD simulations were also run to investigate the formation of coiled nanowires (movie S2 and fig. S9), demonstrating that a decrease of surface energy drives the coiling process.

We prepared strontium-POM (Sr-POM) nanowires through the same method (fig. S10), and the structures were similar to that of Ca-POM nanowires (fig. S11 and table S1). We also tried many other metal cations and POMs; however, no nanowires were obtained when Ca^{2+}/Sr^{2+} or PTA was replaced (SM materials and methods 1.5 and figs. S12 to S16). The concentration,

solvent, and properties of counterions and POM anions influence the coordination modes of counterions and POM anions and their collective states (*12*, *13*). The formation of nanowires requires not only the 1D coordination mode of counterions and POM anions but also the nucleation of building blocks of nanowires from the reaction system and their connection into stable nanowire structures. For other metal counterions and POM anions, they show different self-assembly behavior with Ca^{2+} or Sr^{2+} and PTA anions. The mixed solvents and surfactants used in this work also influence their coordination modes and assembly behavior. Thus, not all metal counterions and POM anions can be prepared into nanowires in this experimental condition. The ratio of $Ca(NO_3)_2$ to PTA added into the reaction also influenced the products (fig. S17). MD simulation results showed that when there was excessive $PW_{12}O_{40}^{3-}$, they would combine more protonated oleyamine, and the nanoclusters tend to be aggregate and form nanosheets (fig. S18A). When there was excessive Ca^{2+}, it would combine with oleylamine, and some nanoparticles formed (fig. S18B), which is consistent with experimental results.

Experimentally, we found that the nanowires dispersion readily formed a gel (SM materials and methods 1.4), and the Tyndall effect is shown in Fig. 2A, inset. Inversion of a test tube showed that the gel was stable, and after standing for several hours at room temperature, a freestanding gel could be obtained (Fig. 2A). The critical gel concentration (CGC) was 0.53%, denoting the minimum mass percent of nanowires necessary for gelation of organic liquids. As shown in Fig. 2, B and C, nanowires intersected with each other in junctions and formed 3D networks in the gel. As shown in Fig. 2D, nanowires outside junctions were parallel to each other and formed a layer. Interwoven nanowires and parallel nanowires formed the structure with many cavities, which could lock organic liquids. For wet gels, the existence of octane in the gel showed plump vesicles (fig. S19A), whereas the vesicles were shriveled after the gel was dried (fig. S19B), and the originally convex morphology changed to the concave morphology. We prepared various organogels with low CGCs (Fig. 2E, fig. S20, and table S2), demonstrating that nanowires were widely applicable gelators for organic liquids. Sr-POM nanowires could also be used to prepare organogels (fig. S21); however, for the same organic liquid, the CGC of Sr-POM nanowires was higher than that of Ca-POM nanowires (table S3) because of the higher relative atomic mass of Sr. Moreover, through removing solvents in gels by means of distillation and centrifugation, nanowires in gels could be recycled more than 10 times to prepare gels (SM materials and methods 1.7 and

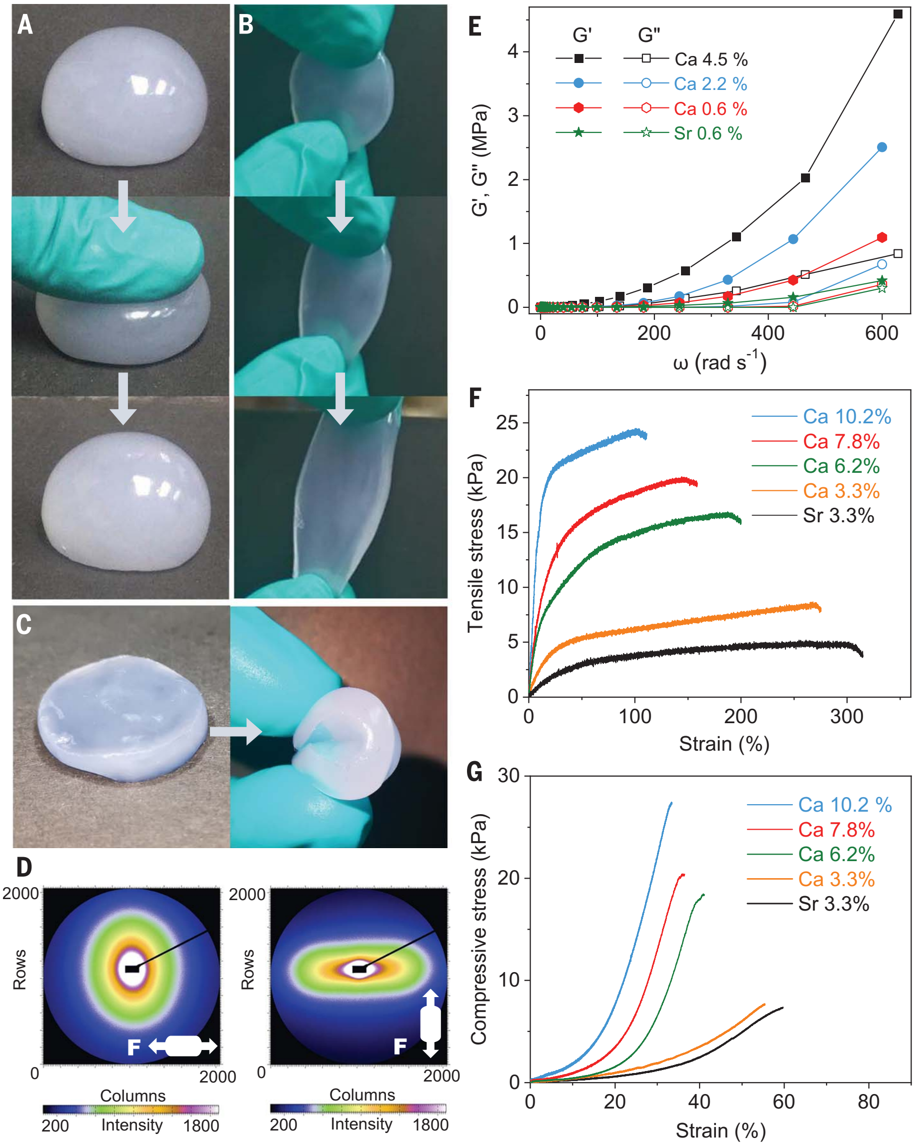


Fig. 3. Mechanical behaviors of nanowires-octane gels. (**A**) Photographs of the gel, which was compressed. (**B**) Photographs of the gel flake, which was stretched. (**C**) Photographs of the gel flake, which was folded. (**D**) SAXS 2D patterns of stretched gel flakes. Tensile directions are shown at bottom right. (**E**) Rheological study of gels in the frequency sweep mode for the strain amplitude of 1%. (**F** and **G**) Typical (F) tensile stress-strain curves and (G) compressive stress-strain curves of gels. Element symbols and numbers in the color keys in (E) to (G) indicate the type of nanowires and their mass percent in gels. For example, Ca 10.2% indicates Ca-POM nanowire-octane gels with 10.2% nanowires.

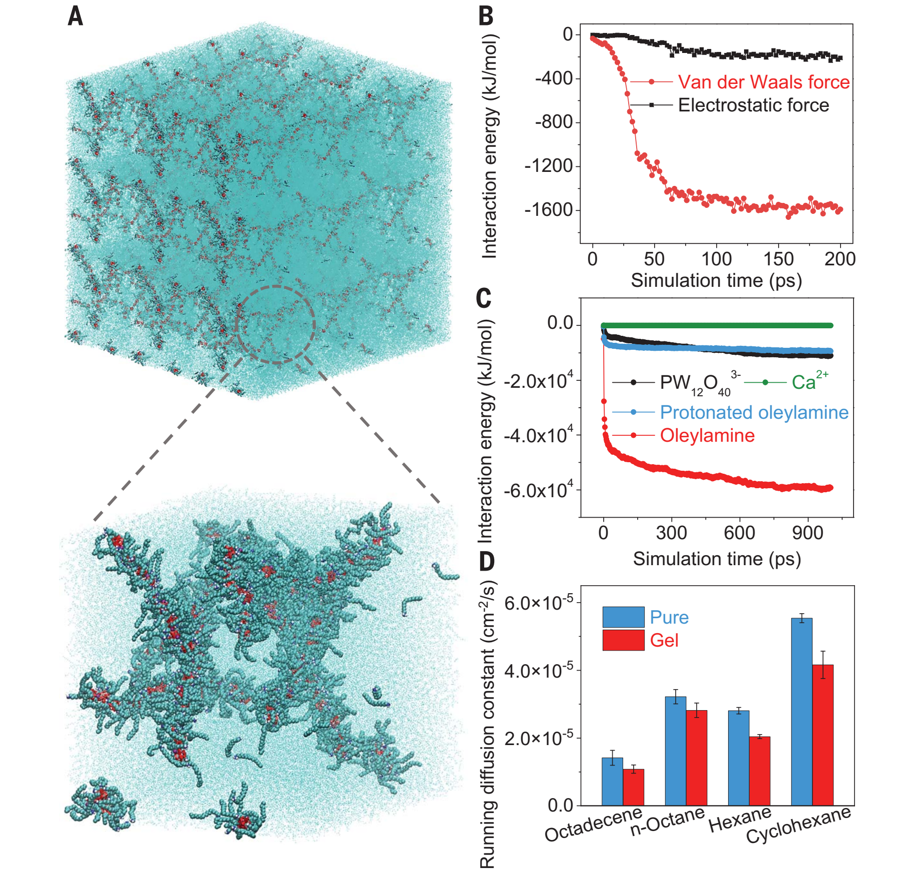


Fig. 4. MD simulations of Ca-POM nanowires-based gels. (**A**) Schematic diagram of the nanowire-octane gel. Small blue models indicate octane, and large blue-red models indicate nanowires. (**B**) The interaction energy between nanowires. (**C**) The interaction energies between octane and main parts of nanowires. (**D**) The running diffusion constants of some pure organic liquids and these liquids trapped in nanowire networks.

fig. S22). The purity of recovered organic liquids from gels was tested as higher than 99.5% (fig. S23), with a little ethanol impurity introduced during the washing process. Nanowires-based gels were stable, and no observable change occurred after 2 months of storage in sealed containers (fig. S24). The gel also remained stable after being frozen with liquid nitrogen (Fig. 2F). The production yield of nanowires could be increased through increasing the quantity of raw materials with equal proportion (SM materials and methods 1.3). As shown in Fig. 2G and fig. S25, gels with bigger sizes were obtained in two batches, with 603 and 373 g octane trapped, respectively. As shown in Fig. 2H, the spilled petrol (~330 ml) on water could form a gel with nanowires, and the freestanding nanowires-petrol gel could be overall removed from water. Then petrol could be separated from the gel through distillation (SM materials and methods 1.8).

The gel was elastic and flexible (Fig. 3, A to C). It could quickly return to its original shape after being compressed (movie S3), and when Ca-POM nanowires in the gel were ~60%, the gel could bounce (movie S4). When the size of gels greatly increased, its elasticity was maintained (movie S5). As shown in small-angle x-ray scattering (SAXS) results (Fig. 3D), when the gel flake was stretched, the SAXS 2D pattern was anisotropic, demonstrating that nanowires in the gel were oriented to a certain extent under tension.

As shown in Fig. 3E, storage modulus (G′) was higher than loss modulus (G″) for a gel, indicating that the gel exhibited more elastic property than viscosity. Because there is a 3D nanowire network in the gels, and because in the low angular velocity (ω) range G′ was higher than G″, there must be a physical cross-link in gels that results from entangling and interactions between nanowires. When nanowires in gels increased, G′ increased far more than G″, with a greater increase in the elasticity, because the density of cross-linked nanowire networks became higher. With ω increasing, nanowires in the gel were partly oriented under external force, so G′ increased rapidly. Sr is heavier than Ca, so there were fewer nanowires in the Sr-POM nanowires-octane gel than in the Ca-POM nanowires-octane gel with the same content of nanowires; thus, G′ of the gel with 0.6% Sr-POM nanowires was lower than the gel with 0.6% Ca-POM nanowires. As shown in Fig. 3F, the stress initially increased quickly and linearly with strain, mainly induced by elastic deformation of nanowire networks. After the yield point, nanowire networks gradually broke under tension, and deformation of gels was permanent. Last, a crack appeared on the gel that propagated under tension until the gel broke. As shown in Fig. 3G, the compressive stress initially increased slowly and linearly with the strain, mainly induced by elastic deformation of the nanowire networks. The stress subsequently increased faster with the strain, induced by permanent deformation of the nanowire networks. In this stage, the energy dissipation involved dissociation of the nanowire networks (*14*). Cracks on the gel gradually increased, causing the gel to be broken into several large pieces, rather than a catastrophic fracture into many small pieces. As shown in tables S4 and S5, when Ca-POM nanowires in gels increased from 3.3 to 10.2%, the tensile elastic modulus, tensile strength, compressive elastic modulus, and compressive fracture strength of the gels all increased, but the elongation at break and the compressive fracture strain decreased, which was also induced by the increase of density of the cross-linked nanowire networks. The octane content affected both the free volume and the lubricity among nanowires, so when the octane content was higher, deformability of gels was better (*15*). For the same gel, the tensile elastic modulus was higher than the compressive elastic modulus (tables S4 and S5). When the gel was stretched, nanowires were oriented to a certain extent under tension (Fig. 3D), and while the gel was compressed, there was no alignment of nanowires to resist compression. The mechanical properties of gels with 3.3% Ca-POM nanowires were better than that of gels with 3.3% Sr-POM nanowires, further demonstrating that the Ca-POM nanowires were better than Sr-POM nanowires as gelators. As shown in fig. S26, the strength of the gel improved when the standing time increased from 1 to 2 days, but further improvement was not apparent after 2 days. Nanowire-based gels were obtained through standing nanowire dispersions for several hours, achieving a preliminary stability of the nanowire networks. Through further standing, nanowire networks and trapped liquids achieved equilibrium. At this point, the gel was fully “mature” and did not evolve further (*16*). Organogels with previously reported high solvent content were usually not freestanding, and most of them did not perform tensile and compression tests. As shown in fig. S27, G′ of the nanowire-based gel was higher than most of the organogels with previously reported high solvents

content, demonstrating the good elasticity of the nanowire-based gel.

We used MD simulations to investigate Ca-POM nanowire-based gels. First, nine nanowires and 24,850 octane molecules were put into a closed system with the dimensions of 20 by 20 by 20 nm. After running MD simulations, nanowires formed networks driven by a decrease of potential energy (Fig. 4A). In addition to van der Waals forces, there were also electrostatic forces between nanowires (Fig. 4B), from the interaction between Ca^{2+} and $PW_{12}O_{40}{}^{3-}$ nanoclusters. The interactions between nanowires and octane, hexane, cyclohexane, and octadecene were strong, whereas the interactions between nanowires and chloroform were weak (fig. S28), so the nanowires dispersion in chloroform could not form a freestanding gel, and only the gel in a tube was obtained. As shown in Fig. 4C, the octane interacted with all of the main parts of the nanowires, and the interaction between octane and oleylamine was the strongest. As shown in Fig. 4D, the running diffusion constants of octane, hexane, cyclohexane, and octadecene in the pure state were all higher than the running diffusion constants of them in the gel, respectively, demonstrating that diffusion of organic liquids was limited by nanowire networks (SM materials and methods 3.4) (*17*, *18*). Thus, the good mechanical properties of gels were derived not only from entanglement of nanowires and multilevel interactions between nanowires contributed by electrostatics force and van der Waals force but also by strong interactions between nanowires and organic liquids.

We developed a facile room-temperature synthesis method of AE-POM nanowires. These nanowires could form 3D networks through physical cross-link in the dispersion. Through simply stirring and standing, more than 10 kinds of organic liquids could be locked, such as petrol and octane, and freestanding and elastic organogels were obtained without any additives. On the basis of this method, we easily achieved scalable batch production of nanowires, and kilogram-scale organic liquids could be trapped. Nanowires in gels could be recycled more than 10 times through distillation and centrifugation. And nanowire-based organogels were stable even at liquid-nitrogen temperature. These materials enable semisolidification and spill recovery of organic liquids.

REFERENCES AND NOTES

1. Y. S. Zhang, A. Khademhosseini, *Science* **356**, eaaf3627 (2017).
2. J. A. Burdick, W. L. Murphy, *Nat. Commun.* **3**, 1269 (2012).
3. E. M. Ahmed, *J. Adv. Res.* **6**, 105–121 (2015).
4. S. M. Zhang, X. Wang, *ACS Mater. Lett.* **2**, 639–643 (2020).
5. S. Zhang *et al*., *J. Am. Chem. Soc.* **142**, 1375–1381 (2020).
6. S. Hu, H. Liu, P. Wang, X. Wang, *J. Am. Chem. Soc.* **135**, 11115–11124 (2013).
7. S. M. Zhang *et al*., *Adv. Funct. Mater.* **31**, 2100703 (2021).
8. X. Xu *et al*., *Small* **12**, 2982–2990 (2016).
9. J. E. Boulicault, S. Alves, R. B. Cole, *J. Am. Soc. Mass Spectrom.* **27**, 1301–1313 (2016).
10. R. Rakhshaee, Y. Noorani, *Adv. Powder Technol.* **28**, 1797–1814 (2017).
11. P. Hudec, K. Prandova, *Collect. Czech. Chem. Commun.* **60**, 443–450 (1995).
12. A. Chaumont, G. Wipff, *Phys. Chem. Chem. Phys.* **10**, 6940–6953 (2008).
13. J. M. Maestre, X. Lopez, C. Bo, J. M. Poblet, N. Casañ-Pastor, *J. Am. Chem. Soc.* **123**, 3749–3758 (2001).
14. J. K. Hao, R. A. Weiss, *Polymer* **54**, 2174–2182 (2013).
15. H. H. Hariri, A. M. Lehaf, J. B. Schlenoff, *Macromolecules* **45**, 9364–9372 (2012).
16. V. Normand, D. L. Lootens, E. Amici, K. P. Plucknett, P. Aymard, *Biomacromolecules* **1**, 730–738 (2000).
17. O. Bénichou, P. Illien, G. Oshanin, A. Sarracino, R. Voituriez, *Phys. Rev. Lett.* **115**, 220601 (2015).
18. A. M. Berezhkovskii, L. Dagdug, S. M. Bezrukov, *Biophys. J.* **106**, L09–L11 (2014).

ACKNOWLEDGMENTS

We are grateful for the support of Shuimu Tsinghua Scholar Program and XPLORER PRIZE. **Funding:** This work was supported by the National Natural Science Foundation of China (22035004), the National Key R&D Program of China (2017YFA0700101), and a China Postdoctoral Science Foundation–funded project (2020TQ0164). **Author contributions:** X.W. initiated and guided the research. S.Z. designed and performed the experiments and wrote and revised the manuscript. W.S. performed the simulations. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm7574
Materials and Methods
Figs. S1 to S28
Tables S1 to S5
References (*19*–*37*)
Movies S1 to S5

Submitted 9 October 2021; resubmitted 11 January 2022
Accepted 1 June 2022
10.1126/science.abm7574

EVOLUTION

Pleiotropic effects of trans-regulatory mutations on fitness and gene expression

Pétra Vande Zande[1]†, Mark S. Hill[2]‡, Patricia J. Wittkopp[1,2]*

Variation in gene expression arises from cis- and trans-regulatory mutations, which contribute differentially to expression divergence. We compare the impacts on gene expression and fitness resulting from cis- and trans-regulatory mutations in *Saccharomyces cerevisiae*, with a focus on the *TDH3* gene. We use the effects of cis-regulatory mutations to infer effects of trans-regulatory mutations attributable to impacts beyond the focal gene, revealing a distribution of pleiotropic effects. Cis- and trans-regulatory mutations had different effects on gene expression with pleiotropic effects of trans-regulatory mutants affecting expression of genes both in parallel to and downstream of the focal gene. The more widespread and deleterious effects of trans-regulatory mutations we observed are consistent with their decreasing relative contribution to expression differences over evolutionary time.

Heritable variation in gene expression is widespread within and between species and often contributes to phenotypic diversity (*1*). This variation arises from mutations that alter activity of the regulatory networks that control gene expression. Each regulatory mutation can act in cis or in trans with respect to a specific gene. Cis-regulatory mutations tend to be located close to the focal gene and often affect noncoding sequences that regulate the focal gene's expression such as promoters or enhancers. Trans-regulatory mutations can be located anywhere in the genome and can affect either coding or noncoding sequences of genes that influence the focal gene's expression through activity of a diffusible molecule such as a protein or RNA. Prior work has shown that trans-regulatory variants appear to be the primary source of mRNA expression differences within a species but the relative contribution of cis-regulatory variants often increases with evolutionary time (*2*–*6*). Understanding how and why these classes of regulatory mutations contribute differently to variation in gene expression is important for understanding how gene expression evolves.

Differences in the way cis- and trans-regulatory mutations affect gene expression might contribute to a preferential fixation of cis-regulatory variants relative to trans (*6*–*9*). A cis-regulatory mutation alters expression of a focal gene, which can in turn have effects on expression of downstream genes (orange box in Fig. 1). By contrast, a trans-regulatory mutation affecting expression of the same focal gene might have effects comparable to the cis-regulatory mutation plus independent effects on expression of other genes, each with its own downstream consequences (blue box in Fig. 1). Trans-regulatory mutations are thus predicted to have more widespread effects on gene expression than mutations altering expression of the focal gene in cis. Consequently, trans-regulatory mutations might be fixed less often than cis-regulatory mutations because mutations that are more pleiotropic (i.e., affect more traits) are predicted to be more deleterious (*10*). This hypothesis has been difficult to test directly, however, because it is notoriously difficult to disentangle the effects of a pleiotropic mutation on one trait from its effects on others (*11*–*13*).

We examine the pleiotropic effects of trans-regulatory mutations by using cis-regulatory mutations to separate the effects of a trans-regulatory mutation caused by its influence on a focal gene from its effects caused by impacts on other genes. We separate these mutational effects for fitness and gene expression by measuring relative growth rate and expression profiles (using RNA-seq) for strains of the baker's yeast *Saccharomyces cerevisiae* carrying either a cis- or trans-regulatory mutation affecting expression of the focal gene. We focus on mutations affecting expression of the *TDH3* gene, which encodes a glyceraldehyde-3-phosphate dehydrogenase (GAPDH), because prior studies have systematically identified and isolated individual cis- and trans-regulatory mutations that affect its expression (*14*–*16*). To the best of our knowledge, comparable data are not available for other genes in *S. cerevisiae* or other eukaryotic species.

We analyzed the effects of five cis-regulatory mutants on fitness and gene expression that caused expression of *TDH3* to vary from 0 to ~135% of wild type (WT) levels (data S1). These mutants were selected to cover the range of effects observed previously for 348 mutant alleles of the 678-bp (base pair) *TDH3* promoter (*14*, *17*, *18*). The five cis-regulatory mutants were genetically identical except that the 0% *TDH3* expression strain carried a deletion of the *TDH3* gene, the 20, 50, and 85% *TDH3* expression strains each carried a single point mutation in the *TDH3* promoter, and the 135%

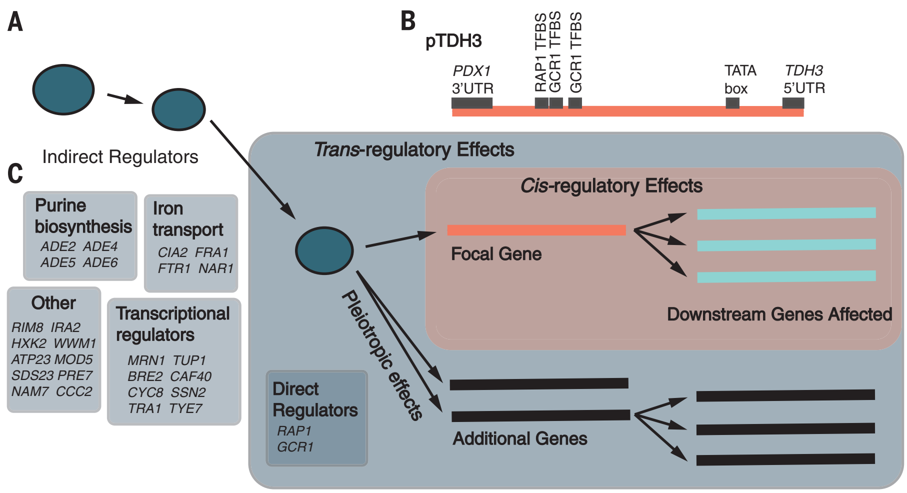


Fig. 1. Cis- and trans-regulatory mutations have different effects on gene expression. (**A**) Trans-regulatory mutations (blue), in either indirect or direct regulators, influence expression of a focal gene (orange), which in turn influences expression of downstream genes (light blue), either directly or indirectly, and can also influence additional genes (black). (**B**) Schematic showing the 678-bp cis-regulatory sequence (promoter) for the *S. cerevisiae TDH3* gene (*pTDH3*) used as a focal gene for this work (see fig. S1 and data S1). (**C**) Previously identified (*16*) trans-regulators of *TDH3* expression harboring mutations tested in this work (see fig. S1 for detail).

[1]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, USA. [2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA.
*Corresponding author. Email: wittkopp@umich.edu
†Current address: Department of Microbiology and Immunology, University of Minnesota, Minneapolis, Minnesota, USA.
‡Current address: Cancer Evolution and Genome Instability Laboratory, University College London Cancer Institute and The Francis Crick Institute, London, UK.

TDH3 expression strain carried a duplication of the *TDH3* allele harboring a point mutation with a copy of URA3 separating the two copies of *TDH3* (fig. S1). The impact of each of these mutations on gene expression was determined by comparing expression in these strains to expression in the progenitor strain lacking any mutations in *TDH3*, either without (for the 0, 20, 50, and 85% strains) or with (for the 135% strain) the extra *URA3* gene (fig. S1, see methods). Simultaneously, we analyzed the effects of 35 trans-regulatory mutant strains carrying mutations that caused *TDH3* expression to vary from ~6 to ~130% of WT levels (data S1). Of these 35 strains, 26 carried single mutations with a range of effects similar to those captured by the five cis-regulatory mutants as well as those observed for >1400 trans-regulatory mutations introduced randomly by ethyl methanesulfonate (*15*). The remaining nine strains each carried one to six mutations in either *RAP1* and *GCR1*, which directly regulate *TDH3* (*16*) (fig. S1 and data S1).

To separate the fitness effects of trans-regulatory mutations attributable to changes in *TDH3* expression from the fitness effects attributable to the pleiotropic impacts of these mutations on other genes, we first defined the relationship between *TDH3* expression and fitness using only the cis-regulatory mutants. Relative fitness was estimated for each mutant based on measures of growth rate under the same conditions used to grow cells for expression profiling (see methods). To predict the fitness effects of any change in *TDH3* expression between 0 and 135% of WT expression, we fit a local polynomial regression (LOESS) curve to these data. The continuous relationship between fitness and *TDH3* expression level observed (Fig. 2A) is consistent with prior work using competitive growth to estimate fitness of 43 cis-regulatory mutations in *TDH3* (*17*, *18*).

Using these inferred effects of changes in *TDH3* expression on fitness, we estimated the pleiotropic fitness effects of each trans-regulatory mutant by comparing its measured fitness to the fitness predicted for a cis-regulatory mutant with the same change in *TDH3* expression (Fig. 2B). Excluding two flocculant trans-regulatory mutants for unreliable estimates of growth rate, 52% of mutants (17 of 33) had significant deleterious pleiotropic effects based on the LOESS regression curve falling above their 99% confidence intervals (CI) for fitness. Only one mutant had a significant beneficial pleiotropic effect; this mutant had a mutation in *IRA2*, which has also been found to harbor beneficial mutations in other studies (*19*–*21*). The remaining 15 trans-regulatory mutants (45%) showed fitness effects comparable to cis-regulatory mutants with similar impacts on *TDH3*. Overall, this empirical distribution of pleiotropic fitness effects was skewed toward deleterious effects (Fig. 2C). Mutations with the largest deleterious fitness effects tended to also cause the largest decreases in *TDH3* expression, although some mutations with large deleterious effects had little effect on *TDH3* expression (Fig. 2C, insert). An additional 1106 gene deletions that affected *TDH3* expression in trans showed a similar distribution of pleiotropic effects (fig. S2) (*22*, *23*)

To test the idea that trans-regulatory mutations are more likely to be deleterious because they tend to be more pleiotropic, we compared the number of genes considered significantly differentially expressed in the cis- and trans-regulatory mutants at a false discovery rate (FDR) of 10%. Overall, we found no statistically significant difference in the median number of differentially expressed genes between the five cis- and 35 trans-regulatory mutants (Fig. 2D, permutation test *P* value: 0.14; fig. S3A) but observed significantly more variable effects for the trans-regulatory mutants (permutation test *P* value = 0.01; fig. S3B). Although the power of this test was limited by the small number of cis-regulatory mutants included, the absence of a larger median effect of trans-regulatory mutants might be due to differences in the severity of mutational effects between the sets of cis- and trans-regulatory mutants examined. To test this possibility, we examined the effects of cis- and trans-regulatory mutants on the number of differentially expressed genes while taking their effect on *TDH3* expression into account. We found that most trans-regulatory mutants showed more widespread effects on gene expression than cis-regulatory mutants with the same effect on *TDH3* expression (Fig. 2E).

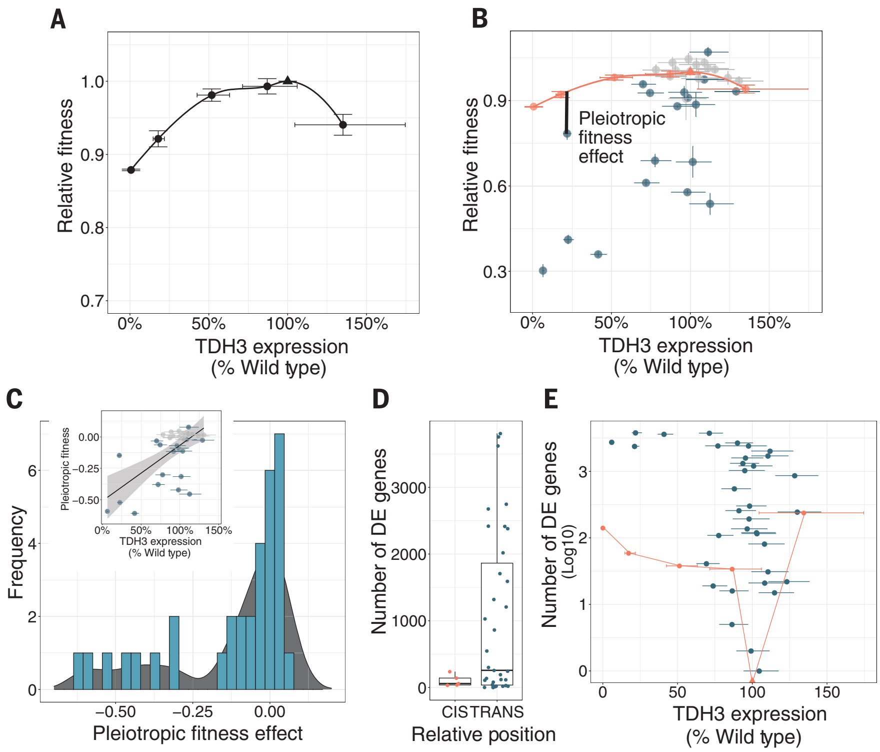


Fig. 2. Pleiotropic effects of trans-regulatory mutations on fitness and gene expression. (**A**) Relative fitness is shown for cis-regulatory mutants and a wild-type strain (triangle) with different levels of *TDH3* expression. Line fitted with a local polynomial regression (LOESS). (**B**) Relative fitness and *TDH3* expression of trans-regulatory mutants is shown, with the cis-regulatory mutants and fitted LOESS curve from (A) in orange. Pleiotropic fitness effect of one trans-regulatory mutant is indicated by the solid black line; mutants with significant pleiotropic fitness effects are blue and those that are not significant are gray. (**C**) Histogram summarizing pleiotropic fitness effects of all trans-regulatory mutants. A smoothed density distribution derived from this histogram is underlaid in black. Inset shows the same pleiotropic fitness effects plotted versus the effect on *TDH3* expression including a robust linear regression (black line). Gray shaded area, 95% CI. (**D**) Number of significantly differentially expressed (DE) genes at a 10% FDR, shown for cis-regulatory mutants (orange) and trans-regulatory mutants (blue). Boxplots show median and quartile values. (**E**) The $\log_{10}$ number of significantly differentially expressed genes at a 10% FDR is shown for each mutant, plotted according to the mutant's impact on *TDH3* expression. Cis-regulatory mutants (circles) and wild-type strain (triangle) are connected by straight line segments. In all panels, error bars represent one standard error of the mean.

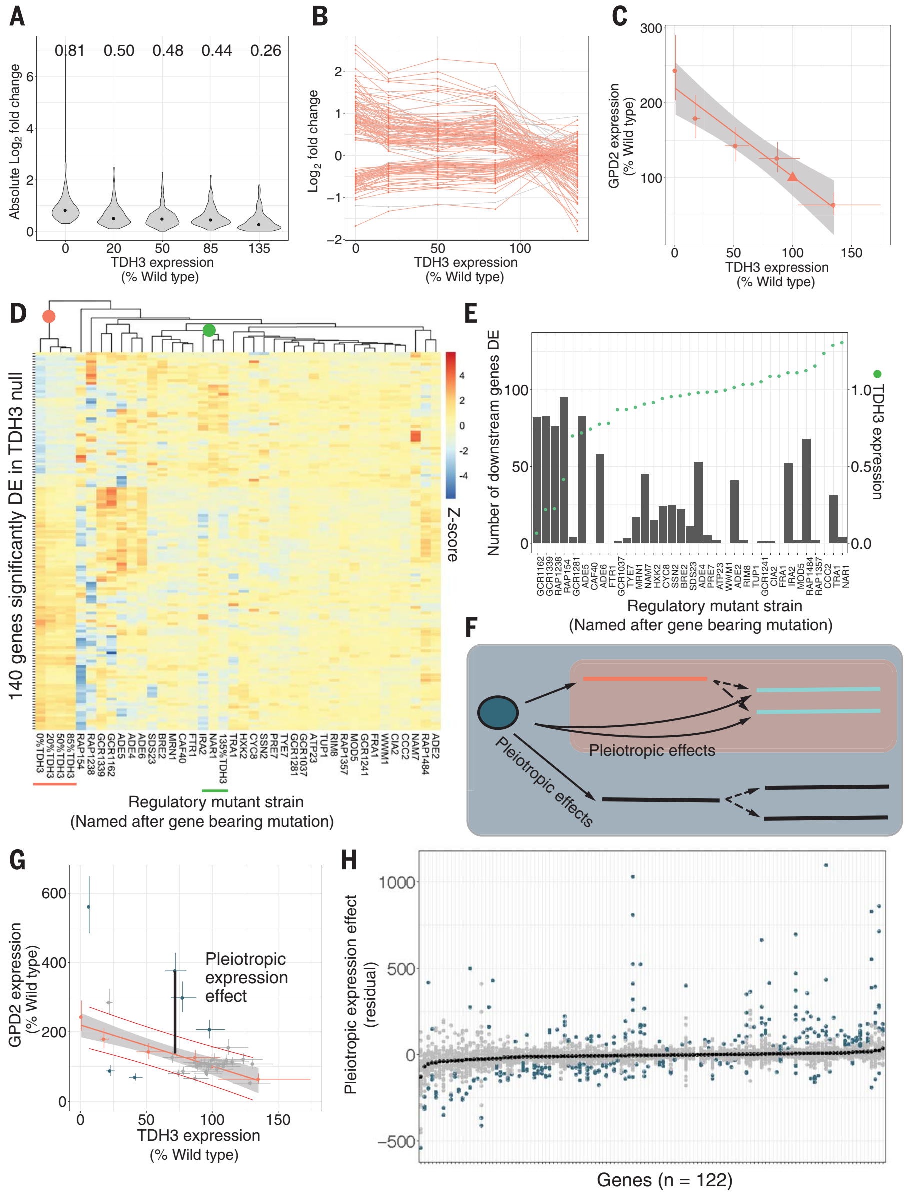


Fig. 3. cis- and trans-regulatory mutants have distinct effects on expression of genes downstream of *TDH3*. (**A**) Violin plots showing absolute $\log_2$ fold changes in the cis-regulatory mutants for the 140 genes identified as downstream of *TDH3*, as they were significantly differentially expressed (DE) in the *TDH3* null mutant. Median absolute $\log_2$ fold changes are shown above each plot and indicated with black dots. (**B**) Log_2 fold changes in cis-regulatory mutants are shown for the same 140 genes downstream of *TDH3* connected by line segments. Genes whose expression was not significantly linearly correlated with *TDH3* expression are shown in gray. (**C**) Expression of *TDH3* and *GPD2* is shown for the cis-regulatory mutants (circles) and the WT strain (triangle). The best fit linear regression line and 95% CI (gray shaded area) are shown. (**D**) Heatmap of the 140 genes downstream of *TDH3* showing $\log_2$ fold changes in all cis- and trans-regulatory mutants; genes are rows and mutants—named after the gene bearing the mutation in that mutant—are columns. Color intensity is scaled by row (by gene) and represents z scores. Mutants are hierarchically clustered as shown by the dendrogram. (**E**) The number of downstream genes that are also significantly differentially expressed in each trans-regulatory mutant is shown as a column relating to the left *y* axis. The expression level of *TDH3* in that mutant is shown as a green point relating to the right *y* axis. (**F**) Schematic shows that trans-regulatory mutants can have pleiotropic effects on genes downstream of *TDH3* not mediated by their effect on *TDH3* as well as pleiotropic effects on genes in parallel to *TDH3*. (**G**) Expression of *GPD2* and *TDH3* is shown for the trans-regulatory mutants, with the expression and linear regression from cis-regulatory mutants from (C) included in orange. Red lines delineate 95% prediction intervals for the cis-regulatory mutant relationship between *TDH3* and *GPD2*. Effects of trans-regulatory mutants on *GPD2* that are not explained by their effect on *TDH3* (pleiotropic expression effects, example illustrated by black line) are colored blue. (**H**) For each of the 122 genes downstream of *TDH3* with a significant linear relationship to *TDH3* expression in the cis-regulatory mutants (*x* axis), the pleiotropic expression effect (*y* axis), as illustrated in (G), is shown. For each gene, each point represents a different trans-regulatory mutant. Genes are ordered on the *x* axis by median residual (black points). Blue points indicate mutants with significant pleiotropic expression effects, whereas gray points indicate nonsignificant pleiotropic expression effects.

As shown in Fig. 1 trans-regulatory mutants are not only assumed to have effects on expression of genes downstream of the focal gene similar to that of cis-regulatory mutations, but are also assumed have additional pleiotropic effects on expression of other genes in parallel. To test this assumption, we sought to separately analyze the effects of trans-regulatory mutants on expression of genes downstream of and in parallel to *TDH3*. To identify the set of genes downstream of *TDH3*, we identified genes whose expression was significantly altered in the *TDH3* null mutant. We found 140 such genes (10% FDR, *P* value = 0.002, data S3), excluding *TDH3* itself. Of these genes, 49 (35%) were underexpressed in the null mutant relative to the wild-type strain and 91 (65%) were overexpressed (data S3). This gene set was significantly enriched for genes encoding proteins involved in glycolytic processes (fig. S4), suggesting that many expression changes observed in the *TDH3* null mutant might be due to a homeostatic response of the cells to maintain metabolism in the absence of the TDH3p enzymatic activity involved in glycolysis and gluconeogenesis (*24*). These downstream genes were also enriched for genes associated with the gene ontology terms DNA biosynthesis, integration, and transposition (fig. S4), which might result from nonmetabolic functions of TDH3p in processes such as transcriptional silencing and rDNA recombination (*25*).

The median absolute $\log_2$ fold expression changes observed for this set of 140 genes downstream of *TDH3* decreased monotonically as *TDH3* expression approached wild type and was smallest in the cis-regulatory mutant overexpressing *TDH3* (Fig. 3A). Out of these 140 genes, 122 (87%) showed a significant linear relationship with *TDH3* expression in the five cis-regulatory mutants (10% FDR, *P* value = 0.09; fig. S5A), with 44 positively correlated and 78 negatively correlated (Fig. 3B and fig. S5B). For example,

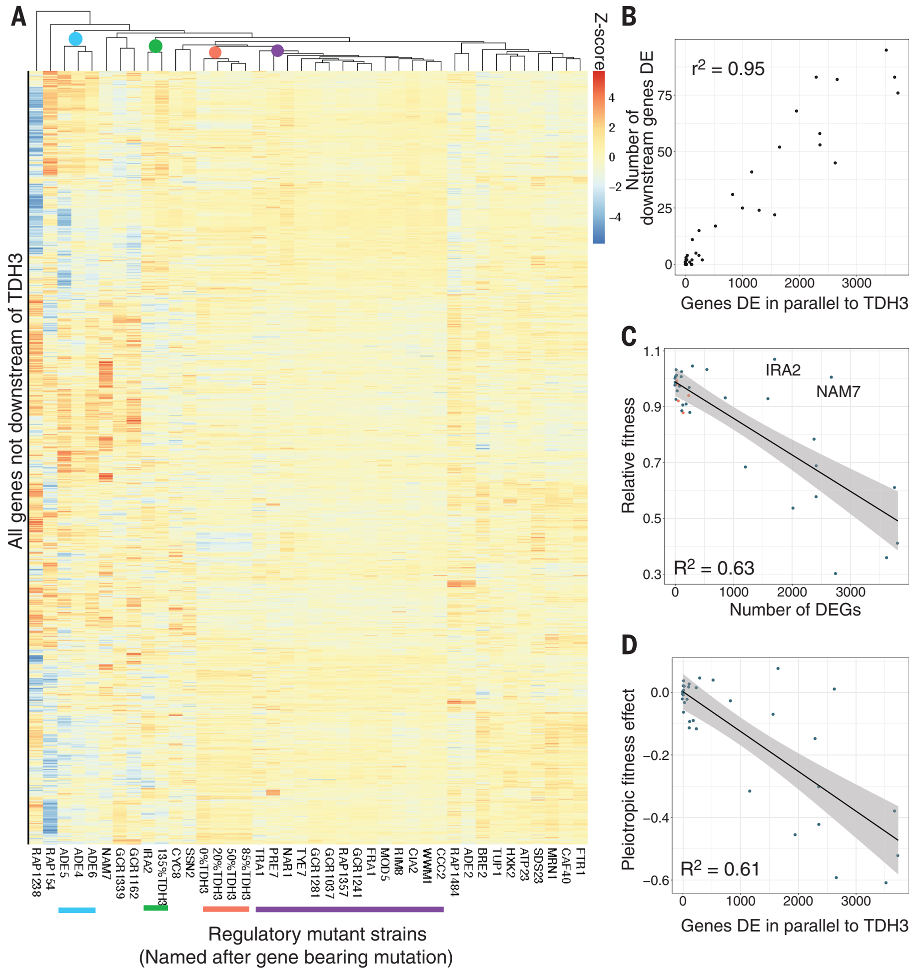


Fig. 4. Larger pleiotropic expression effects of trans-regulatory mutants correlate with negative fitness effects. (**A**) A heatmap shows $\log_2$ fold changes for all genes not downstream of *TDH3* as estimated by DESeq2, in which rows are genes and columns are mutants. Color intensity is scaled by row (by gene) and corresponds to z scores. Hierarchical clustering of mutants is shown by the dendrogram above. Three of the four trans-regulatory mutants bearing mutations in the adenine biosynthesis pathway cluster together and are marked by a blue dot and line. The cis-regulatory mutant overexpressing *TDH3* clusters with *IRA2* (green), whereas the cis-regulatory mutants with reduced *TDH3* expression cluster together (orange). A large cluster of trans-regulatory mutants with small effects across the genome are marked by a purple dot and line. (**B**) For all trans-regulatory mutants, the number of downstream genes differentially expressed is positively correlated with the number of genes differentially expressed in parallel to *TDH3* in that mutant. (**C**) For both cis- (orange) and trans- (blue) regulatory mutants, the total number of differentially expressed genes is a strong predictor of relative fitness. A linear model is shown with gray shading representing 95% CI and two outliers, *IRA2* and *NAM7* are indicated. (**D**) For all trans-regulatory mutants with fitness data (i.e., excluding the two flocculant strains), the pleiotropic fitness effect, as defined in Fig. 2B, is plotted vs the number of genes differentially expressed in parallel to *TDH3* in that trans-regulatory mutant. A linear regression is shown in black with 95% CI in gray.

the *GPD2* gene—which encodes an enzyme two steps away from *TDH3* in the metabolic network—increased linearly when *TDH3* expression was decreased by cis-regulatory mutations (Fig. 3C).

A heatmap showing expression with hierarchical clustering for these 140 genes downstream of *TDH3* in the 40 cis- and trans-regulatory mutants also visually showed the correlation with *TDH3* expression in the cis-regulatory mutants (orange cluster in Fig. 3D). The cis-regulatory mutant overexpressing *TDH3* (135% TDH3) had opposing effects on expression of these genes that caused it to cluster separately, with expression most similar to two trans-regulatory mutants (bearing mutations in *NAR1* and *IRA2*) that also caused overexpression of *TDH3* (green cluster in Fig. 3D). The other 33 trans-regulatory mutants had more distinct patterns of expression for these genes (Fig. 3D), with some mutants (e.g., RAP1484 and IRA2) showing significant changes in expression for many genes despite minimal effects on *TDH3* expression (Fig. 3E). These observations suggest that the pleiotropic effects of trans-regulatory mutants include effects on expression of genes downstream of *TDH3* that are not explained by their effect on *TDH3* (Fig. 3F).

To further explore this possibility, we used the linear models fitted to the cis-regulatory mutant data to predict the change in expression expected for each downstream gene as a result of the effect of the trans-regulatory mutant on *TDH3* expression alone. Deviations from these expectations indicate pleiotropic effects of trans-regulatory mutants on the expression of genes downstream of *TDH3*. For example, 6 of the 35 trans-regulatory mutants showed evidence of a pleiotropic effect on expression of *GPD2*, as indicated by a change in *GPD2* expression further than one standard error outside of the 95% prediction interval for the expression change expected based on cis-regulatory mutants (Fig. 3G). Such pleiotropic effects were observed for at least one trans-regulatory mutant for 111 of the 122 genes downstream of *TDH3*, with the magnitude of the pleiotropic effects (measured as residuals from the gene-specific regression models based on the cis-regulatory mutants) varying among trans-regulatory mutants and genes (Fig. 3H).

These pleiotropic effects on expression of genes downstream of *TDH3* are in addition to the pleiotropic effects of trans-regulatory mutants in parallel to *TDH3*. A heatmap of expression with hierarchical clustering for the 5806 genes not classified as downstream of *TDH3* shows these other effects (Fig. 4A). Of the trans-regulatory mutants, 14 showed minimal effects on expression of these genes (purple cluster in Fig. 4A). The four hypomorphic cis-regulatory mutants clustered together (orange cluster in Fig. 4A) and showed additional effects that appeared to scale with *TDH3* expression level (consistent with fig. S5), suggesting that there might be more genes downstream of *TDH3* than were identified as differentially expressed in the *TDH3* null mutant with the statistical thresholds used. The trans-regulatory mutants with larger effects tended to cluster according to related phenotype or function, such as the two flocculant strains (bearing mutations in *CYC8* and *SSN2*), genes involved in adenine biosynthesis (*ADE4, ADE5,* and *ADE6*), and large-impact mutations in *GCR1*. Finally, the two *RAP1* mutants with the largest impacts on *TDH3* expression (RAP154 and RAP1238) showed the most different expression patterns from the other mutants and each other.

Comparing the impacts of trans-regulatory mutations on gene expression downstream of and in parallel to *TDH3* showed that mutants affecting expression of many genes in parallel to *TDH3* tended to also affect expression of many genes downstream of *TDH3* (Fig. 4B). This observation indicates that some trans-regulatory mutants had larger or smaller overall effects on the transcriptome. However, the impact of a mutation on *TDH3* expression was not a strong predictor for this overall effect: Some mutants with minimal effects on *TDH3* expression altered expression of thousands of other genes whereas other mutants changed *TDH3* expression up to 30% but altered expression of <100 other genes (Fig. 2E). The overall number of genes whose expression was affected by a trans-regulatory mutation was a strong predictor of the relative fitness of the mutant [Fig. 4C; consistent with (*26*)], although two trans-regulatory mutants (*IRA2* and *NAM7*) were notable exceptions (Fig. 4C). Directly comparing the pleiotropic effects of trans-regulatory mutations inferred for fitness and expression of genes in parallel to *TDH3* showed a very similar relationship (Fig. 4D) because the cis-regulatory mutants had much smaller effects on both fitness and expression of other genes than trans-regulatory mutants with similar effects on *TDH3* expression (Fig. 2, B and E).

In summary, we developed a framework to decompose the effects of trans-regulatory mutations attributable to their effect on expression of a focal gene from their effect on other genes in the genome. These data support the widely held but rarely tested assumptions that trans-regulatory mutations tend to be more pleiotropic and more deleterious than cis-regulatory mutations but also provide a more nuanced view of these differences at the level of specific mutations and show that the pleiotropic effects of trans-regulatory mutations might often affect expression of genes both in parallel to and downstream of the focal gene. Studies such as this one that empirically determine and compare the properties of different types of mutations affecting gene expression are critical for understanding how regulatory systems evolve.

REFERENCES AND NOTES

1. D. L. Stern, V. Orgogozo, *Evolution* **62**, 2155–2177 (2008).
2. J. D. Coolon, C. J. McManus, K. R. Stevenson, B. R. Graveley, P. J. Wittkopp, *Genome Res.* **24**, 797–808 (2014).
3. B. P. H. Metzger, P. J. Wittkopp, J. D. Coolon, *Genome Biol. Evol.* **9**, 843–854 (2017).
4. D. Gokhman *et al.*, *Nat. Genet.* **53**, 467–476 (2021).
5. S. A. Signor, S. V. Nuzhdin, *Trends Genet.* **34**, 532–544 (2018).
6. M. S. Hill, P. Vande Zande, P. J. Wittkopp, *Nat. Rev. Genet.* **22**, 203–215 (2021).
7. G. A. Wray, *Nat. Rev. Genet.* **8**, 206–216 (2007).
8. P. J. Wittkopp, B. K. Haerum, A. G. Clark, *Nat. Genet.* **40**, 346–350 (2008).
9. B. Schaefke *et al.*, *Mol. Biol. Evol.* **30**, 2121–2133 (2013).
10. M. Kimura, T. Ohta, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2848–2852 (1974).
11. A. B. Paaby, M. V. Rockman, *Trends Genet.* **29**, 66–73 (2013).
12. J. Zhang, G. P. Wagner, *Trends Genet.* **29**, 383–384 (2013).
13. A. B. Paaby, M. V. Rockman, *Trends Genet.* **29**, 384 (2013).
14. B. P. H. Metzger, D. C. Yuan, J. D. Gruber, F. Duveau, P. J. Wittkopp, *Nature* **521**, 344–347 (2015).
15. B. P. H. Metzger *et al.*, *Mol. Biol. Evol.* **33**, 1131–1146 (2016).
16. F. Duveau *et al.*, *eLife* **10**, e67806 (2021).
17. F. Duveau, W. Toubiana, P. J. Wittkopp, *Mol. Biol. Evol.* **34**, 2908–2912 (2017).
18. F. Duveau *et al.*, *eLife* **7**, e37272 (2018).
19. A. L. Sezmis, M. E. Malerba, D. J. Marshall, M. J. McDonald, *J. Mol. Evol.* **86**, 111–117 (2018).
20. S. Venkataram *et al.*, *Cell* **166**, 1585–1596.e22 (2016).
21. Y. Li *et al.*, *Curr. Biol.* **28**, 515–525.e6 (2018).
22. P. Kemmeren *et al.*, *Cell* **157**, 740–752 (2014).
23. C. J. Maclean *et al.*, *Mol. Biol. Evol.* **34**, 2486–2502 (2017).
24. L. McAlister, M. J. Holland, *J. Biol. Chem.* **260**, 15019–15027 (1985).
25. A. E. Ringel *et al.*, *PLOS Genet.* **9**, e1003871 (2013).
26. D. E. Featherstone, K. Broadie, *BioEssays* **24**, 267–274 (2002).
27. P. V. Zande, M. S. Hill, P. J. Wittkopp, *Zenodo* (2022); https://zenodo.org/record/6567260#.Yp-XlqjMl2w.

ACKNOWLEDGMENTS

We thank B. Metzger, F. Duveau, M. Siddiq, H. Ertl, A. Redgrave, and other members of the Wittkopp laboratory for helpful discussions and feedback on drafts of this manuscript. We also thank B.Z. He (University of Iowa) for sharing the protocol used for RNA-seq. The University of Michigan Advanced Genomics Core, Center for Statistical Consultation and Research, and High-Performance Computing Cluster provided services used to conduct this work. **Funding:** This work was funded by National Institutes of Health grants R35GM118073 and R01GM108826 to P.J.W. and T32GM07544 to P.V.Z. **Author contributions:** P.V.Z. and P.J.W. conceived and designed the study. P.V.Z. performed all experiments, data collection, and data analysis. M.S.H. contributed to the design of the data analysis, especially in the analysis of previously published gene deletion data. P.V.Z. and P.J.W. wrote the manuscript, and M.S.H. provided comments on later drafts. P.V.Z. prepared all figures, code, and supplemental files, with input from P.J.W. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** RNA-seq data is available at GEO accession GSE175398. Code used for data analysis, as well as supporting data files, are available from GitHub at https://github.com/pvz22/Trans-reg_pleiotropy with the version of record at time of publication available at Zenodo (*27*). All other data is included with the manuscript or supplementary material. Mutant strains of *S. cerevisiae* available upon request.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj7185
Materials and Methods
Figs. S1 to S9
References (*28*–*33*)
MDAR Reproducibility Checklist
Data S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

Submitted 28 May 2021; resubmitted 7 February 2022
Accepted 20 May 2022
10.1126/science.abj7185

BIOMEDICINE

Soft, bioresorbable coolers for reversible conduction block of peripheral nerves

Jonathan T. Reeder[1,2,3]†‡*, Zhaoqian Xie[4,5]†, Quansan Yang[3,6]†, Min-Ho Seo[2,3,7]†, Ying Yan[8], Yujun Deng[9], Katherine R. Jinkins[3], Siddharth R. Krishnan[2,3], Claire Liu[3,10], Shannon McKay[10], Emily Patnaude[10], Alexandra Johnson[10], Zichen Zhao[4,5], Moon Joo Kim[11]§, Yameng Xu[12], Ivy Huang[2,3], Raudel Avila[6], Christopher Felicelli[13], Emily Ray[14], Xu Guo[4,5], Wilson Z. Ray[8,14], Yonggang Huang[2,3,6,15], Matthew R. MacEwan[8,14], John A. Rogers[2,3,6,10,16,17,18]*

Implantable devices capable of targeted and reversible blocking of peripheral nerve activity may provide alternatives to opioids for treating pain. Local cooling represents an attractive means for on-demand elimination of pain signals, but traditional technologies are limited by rigid, bulky form factors; imprecise cooling; and requirements for extraction surgeries. Here, we introduce soft, bioresorbable, microfluidic devices that enable delivery of focused, minimally invasive cooling power at arbitrary depths in living tissues with real-time temperature feedback control. Construction with water-soluble, biocompatible materials leads to dissolution and bioresorption as a mechanism to eliminate unnecessary device load and risk to the patient without additional surgeries. Multiweek in vivo trials demonstrate the ability to rapidly and precisely cool peripheral nerves to provide local, on-demand analgesia in rat models for neuropathic pain.

The high therapeutic efficacy of opioids has led to their widespread use despite increasing rates of addiction and opioid-related deaths due to overdose (*1*). The increasing societal burden caused by opiate misuse motivates the development of localized, nonopioid, and nonaddictive pain-management techniques. Miniaturized implantable devices that eliminate pain signals locally in peripheral nerves suggest a potential role for engineering-based treatments that avoid side effects associated with opioids and other analgesics (*2*, *3*). The controlled input of electrical (*4*), pharmacological (*5*), optical (*6*), mechanical (*7*), or thermal (*8*, *9*) stimuli to neural tissue can lead to local and reversible neural blocking.

Of particular interest for the work here is that temporal measures of metabolic, electrogenic, and ionic activity in neural tissue all exhibit a negative temperature dependence (*10*). Local cooling of peripheral nerves decreases conduction velocity and signal amplitude of neural activity (*11*). Blocking of transmission of compound action potentials in mammalian nerves typically occurs below 15°C (*12*), but this threshold can be temporarily increased to near room temperature by a brief heating period preceding the cooling period (*13*, *14*). Cooling applied to peripheral nerves is a promising approach for blocking pain signals because it is nonaddictive, is rapidly reversible, can be applied locally, avoids any onset response, and allows for simultaneous electrical interrogation of the blocked nerve (*15*). Analgesic nerve cooling requires spatiotemporally precise control of temperature to maximize desired outcomes and to minimize the chance of cooling-induced tissue damage (*16*, *17*). Current approaches for nerve cooling rely on rigid, bulky systems that prevent the use of local cooling as a practical approach for peripheral nerve pain management.

An implantable device that provides on-demand local analgesia over a defined timeline followed by subsequent dissolution and bioresorption would represent a qualitative advance in pain-management techniques. We specifically envision clinical use cases for nonopioid management of postoperative acute pain signals in peripheral nerves where (i) aberrant neural signals are well defined in select anatomical regions, (ii) nerves carrying aberrant neural signals are already isolated, and (iii) a need for opioid therapy exists after operation (Fig. 1A). Pain management after amputations, nerve grafts, or spinal decompression surgeries represent examples. Here, the relevant nerves are already isolated and identified; thus, the application of the cuff would be straightforward to integrate into the clinical workflow. In this scheme, implantation of a bioresorbable cooler around the nerve that innervates damaged tissue enables reversible elimination of neural activity and pain signals through the focused application of cooling power (Fig. 1B). Construction from water-soluble materials leads naturally to dissolution of the cooling system after the completion of the healing process and obviates the need for an extraction surgery (Fig. 1C), in a manner conceptually similar to that of other recently reported classes of bioresorbable sensors and therapeutic systems to monitor and accelerate wound healing or recovery processes (*18*, *19*).

We present concepts and device designs for soft, bioresorbable peripheral nerve-cooling and temperature-sensing units that are engineered to reversibly block pain signals with a targeted cooling stimulus over a finite time period matched to patient needs. The technology consists of a hybrid microfluidic and electronic system for cooling and simultaneously measuring the temperature of a peripheral nerve. The elastomeric nature of the microfluidic system and the serpentine shapes of the electrical interconnects yield soft, stretchable mechanics at the device level (Fig. 1D) with effective moduli not substantially higher than those of peripheral nerves [rat sciatic nerve elastic modulus is 0.6 MPa (*20*)]. These electronic and microfluidic systems terminate in a cuff structure with a diameter matched to the rat sciatic nerve (1.5 mm) to provide an intimate mechanical and thermal interface to the nerve, without the need for sutures (Fig. 1E). The curled geometry of the cuff and its elastic nature enables manual unrolling and soft clasping of the nerve. An essential defining characteristic of this system is that it is constructed entirely with water-soluble constituent materials that controllably dissolve to biocompatible end products in the biofluids that are contained in subcutaneous tissues. Figure 1F shows devices wrapped around a silicone phantom nerve and submerged in phosphate-buffered saline (PBS) (pH 7.4) at 75°C, as an accelerated aging test. The results show that the materials largely dissolve within 20 days and that elimination of residues occurs after 50 days under these conditions.

An illustration of the layers of the hybrid microfluidic-electronic device is shown in Fig. 2A. A bioresorbable elastomer, poly(octanediol citrate) (POC), forms the microfluidic system. POC

[1]Knight Campus for Accelerating Scientific Impact, University of Oregon, Eugene, OR, USA. [2]Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA. [3]Querrey Simpson Institute for Bioelectronics, Northwestern University, Evanston, IL, USA. [4]State Key Laboratory of Structural Analysis for Industrial Equipment, Department of Engineering Mechanics, Dalian University of Technology, Dalian, China. [5]Ningbo Institute of Dalian University of Technology, Ningbo, China. [6]Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA. [7]School of Biomedical Convergence Engineering, College of Information and Biomedical Engineering, Pusan National University, Busan, Republic of Korea. [8]Department of Neurological Surgery, Washington University School of Medicine, St. Louis, MO, USA. [9]State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, China. [10]Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA. [11]Department of Chemical Engineering, Northwestern University, Evanston, IL, USA. [12]The Institute of Materials Science and Engineering, Washington University, St. Louis, MO, USA. [13]Department of Pathology, Northwestern University, Evanston, IL, USA. [14]Department of Biomedical Engineering, Washington University, St. Louis, MO, USA. [15]Departments of Civil and Environmental Engineering, Northwestern University, Evanston, IL, USA. [16]Department of Chemistry, Northwestern University, Evanston, IL, USA. [17]Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. [18]Department of Neurological Surgery, Feinberg School of Medicine, Northwestern University, Evanston, IL, USA.
***Corresponding author. Email: jonathanr@science.xyz (J.T.R.); jrogers@northwestern.edu (J.A.R.)**
†These authors contributed equally to this work.
‡Present address: Science Corporation, Alameda, CA, USA.
§Present address: LG Innotek, Seoul, South Korea.

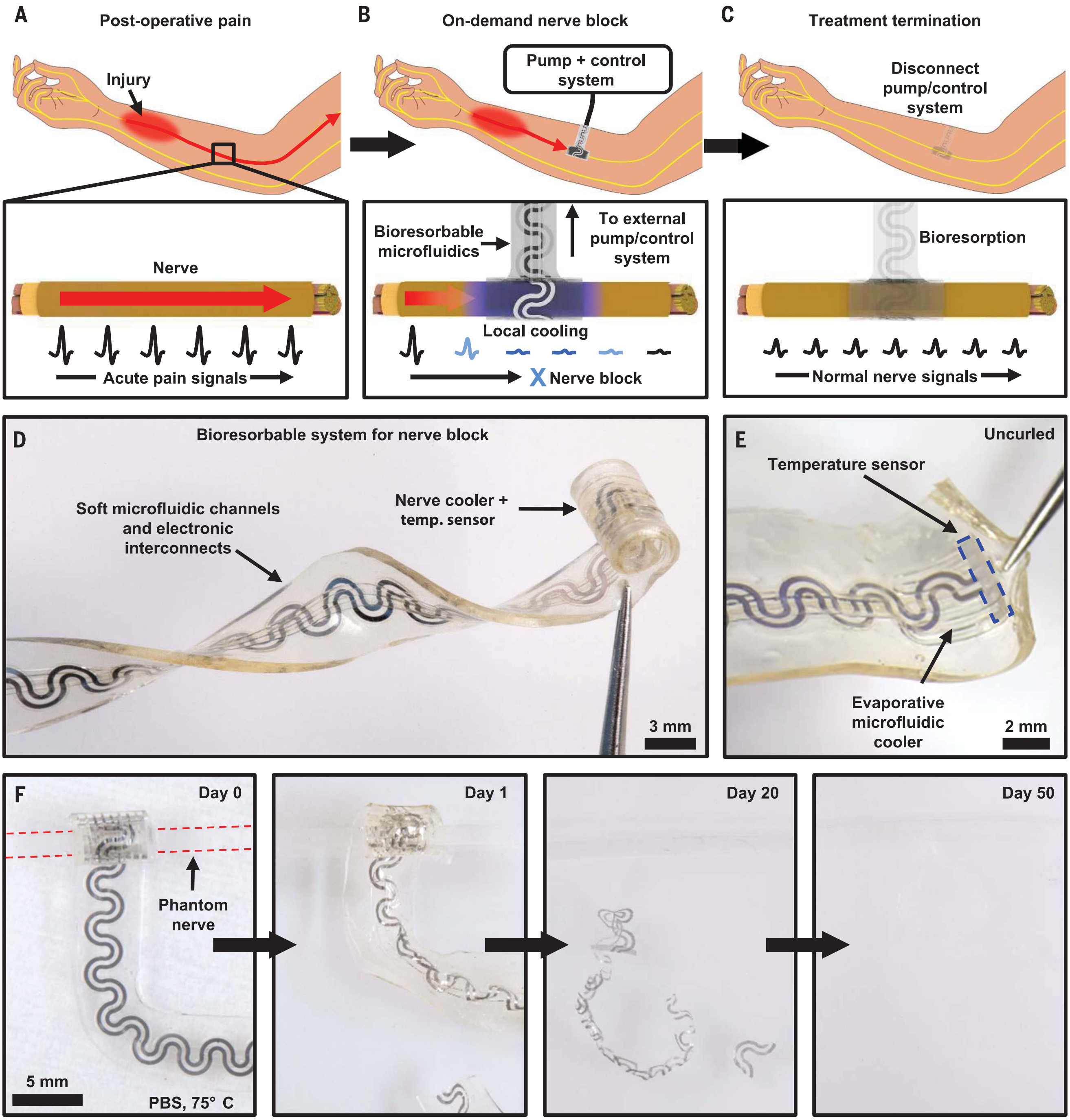


Fig. 1. Soft, bioresorbable, evaporative microfluidic coolers for an on-demand nerve block. (**A**) Transmission of postoperative acute pain signals through peripheral nerves. (**B**) Local cooling of peripheral nerves provides an on-demand nerve block. (**C**) Treatment termination and cooler bioresorption after completion of the healing process. (**D**) A soft, bioresorbable nerve-cooling system that features elastomeric interconnects and a terminal cuff structure. (**E**) A soft, curled, nerve-clasping structure provides secure attachment to nerves without sutures. (**F**) Construction from water-soluble materials enables device dissolution and subsequent bioresorption (PBS at 75°C, pH 7.4).

exhibits an elastic modulus of 2.8 MPa (*21*), controllable rates of degradation by means of surface erosion (*22*), and a demonstrated compatibility with nerves (*23*). Details of the steps for fabrication are provided in fig. S1 and the supplementary materials. The microfluidic system includes transcutaneous colinear interconnects that deliver liquid coolant [perfluoropentane (PFP)] and dry N_2 to a serpentine evaporation chamber (Fig. 2B) in a completely sealed system that provides fluidic access at the ends. The electronic layer lies coplanar with the microfluidic system, with the temperature-sensing element at the distal end of the device. The simultaneous initiation of PFP and N_2 flows into this structure prompts evaporation of PFP at the microfluidic junction between the PFP and N_2 channels and along the serpentine

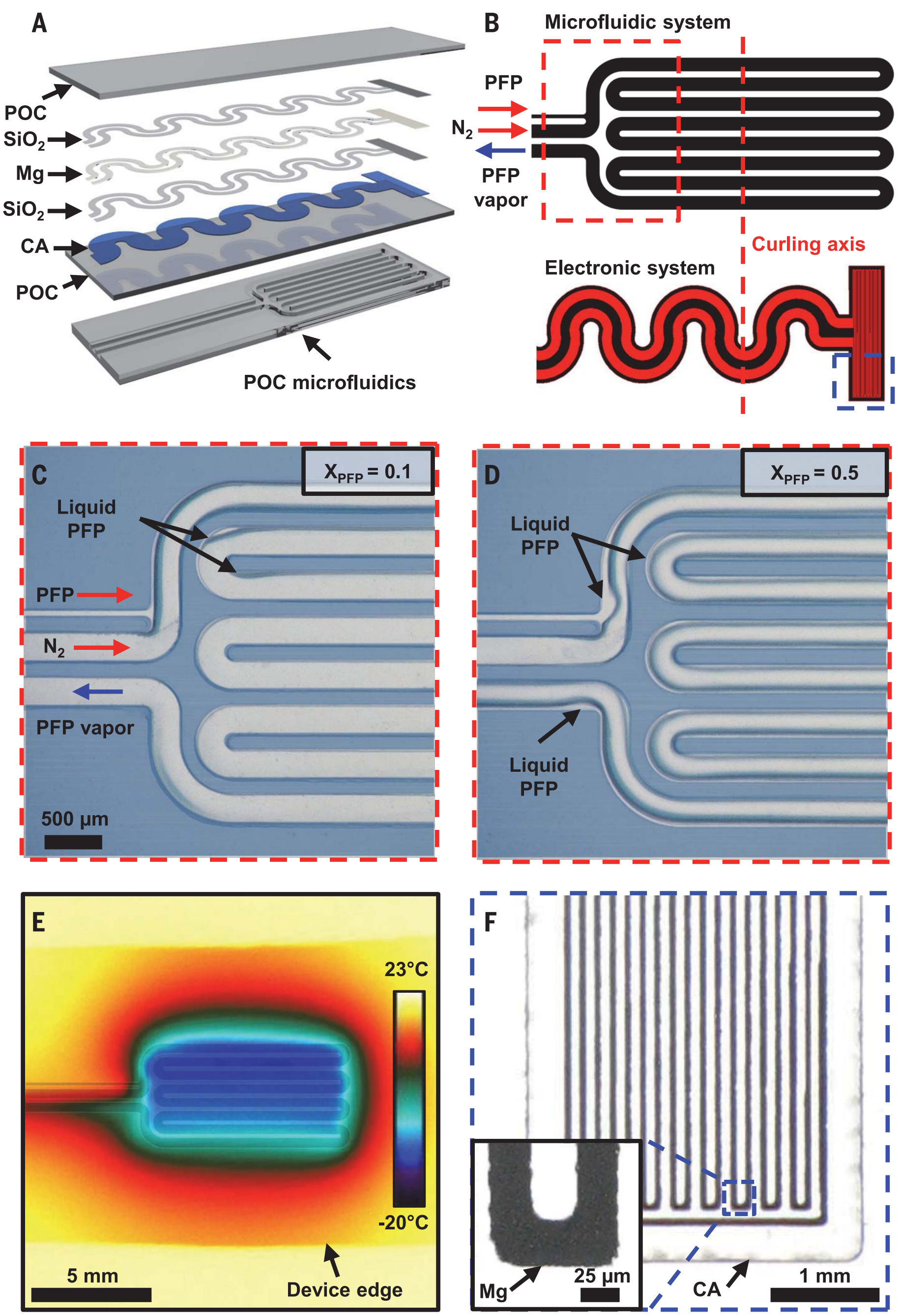


Fig. 2. Evaporative microfluidic cooling and temperature sensing system. (**A**) Exploded device render showing layers of the microfluidic and electronic systems. (**B**) Schematic illustration of the microfluidic and electronic systems design. (**C**) A flow of dry N_2 prompts the complete evaporation of PFP within three serpentine turns at low molar ratio (X_{PFP} = 0.1) inside the serpentine microfluidic channel. (**D**) Increasing the PFP molar fraction (X_{PFP} = 0.5) extends the evaporation throughout the length of the serpentine microfluidic channel. (**E**) Evaporative microfluidic cooling prompts reduction of the temperature of the device surface to −20°C in ambient conditions, as shown by thermal imaging. (**F**) A bioresorbable electronic system for providing nerve-temperature feedback consists of a magnesium resistance-based temperature sensor on a cellulose acetate (CA) substrate.

chamber. PFP, which boils near room temperature (28° to 30°C), is bioinert and compatible with nonfluorinated elastomers. PFP is clinically approved as a propellant in pressurized metered-dose inhalers (*24*), as an intravenous ultrasound contrast agent (*25*), and in therapeutic hypothermia (*26*). The mass flow rates of the PFP and N_2 and the geometry of the evaporation chamber govern the magnitude and localization of the cooling effect. At low PFP molar flow ratios (X_{PFP} = 0.1), the PFP fully evaporates after passing through three serpentines, with marginal liquid PFP buildup at the corners of the microchannels (Fig. 2C). At high molar flow rates (X_{PFP} = 0.5), PFP proceeds through annular flow and passes along the sidewalls of the microchannels (Fig. 2D). This phase change prompts the temperature of a device in a planar, uncurled configuration to drop to −20°C within 2 min after initializing flow in ambient, room-temperature conditions (Fig. 2E). The cooled area of the device is confined predominately to the serpentine evaporation chamber, as governed by the microfluidic channel design and fluid flow rates. A serpentine magnesium trace with a width and length of 25 μm and 72 mm, respectively, provides temperature feedback through the temperature coefficient of resistance of Mg (Fig. 2F). Details for the temperature-sensing system are provided in fig. S2 and the supplementary materials.

Figure 3 summarizes quantitative results of measurements of the efficacy of nerve cooling of a phantom nerve structure and an integrated thermocouple inside a 37°C hydrogel tissue mimic (fig. S3). An enabling property of PFP is that it can be superheated beyond its boiling point, to temperatures as high as 60°C (*27*), because of the energy barrier required for homogeneous nucleation (*28*). This feature prevents premature boiling of the PFP as it passes along the smooth interior of the microfluidic channels to the evaporation chamber. The temperatures of the PFP and N_2 flows equilibrate at 37°C within 90 mm of entering the 37°C environment (fig. S4). A condenser placed distal to the cooling cuff enables recapture of 90% of the evaporated PFP (fig. S5). The delivery of N_2 in a pressure-driven mode mitigates transient pressure spikes that can occur with multiphase microfluidic flows (fig. S6).

Holding the N_2 pressure constant at values ranging from 7 to 90 kPa and sweeping the PFP flow rates from 0 to 900 μl/min yields data that reveal the dependence of the N_2 flow rate on PFP flow rate (Fig. 3A). Experiments based on systematically sweeping the PFP and N_2 flow rates over the same range as in Fig. 3A reveal the effect of molar flow rates on resultant nerve temperature (Fig. 3B). The minimum nerve temperature is −1.4°C, which is achieved at a molar PFP fraction of 0.13 and a N_2 pressure of 90 kPa (Fig. 3C). Additional details regarding conversion to molar flow rates are provided in fig. S7. Figure 3D demonstrates the effect of PFP flow rate on nerve-temperature cooling rate. The maximum cooling rate of 3°C/s occurs with the 300 μl/min case. Arbitrary cooling and rewarming profiles, as governed by the thermal properties of the surrounding media, follow from controlled increments or decrements of the PFP flow rate (Fig. 3E). Experiments indicate sustained

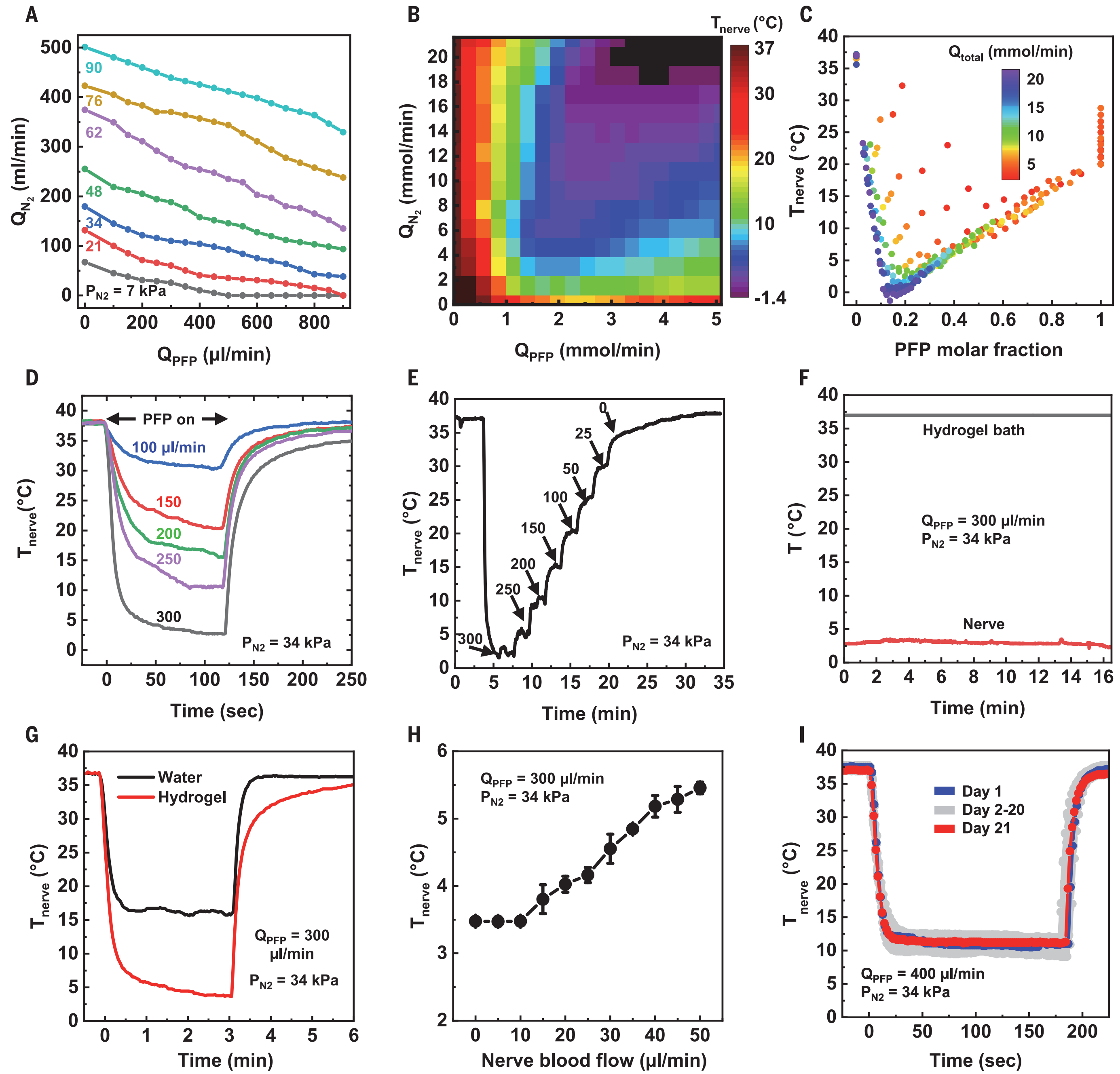


Fig. 3. In vitro studies of nerve-cooling efficacy. (**A**) The N_2 flow rate depends on N_2 pressure (P_{N_2}) and PFP flow rate. Q_{N_2}, N_2 flow rate; Q_{PFP}, PFP flow rate. (**B**) Phantom-nerve temperature depends on PFP and N_2 flow rates. (**C**) A PFP molar fraction of 0.13 produces the lowest nerve temperature (–1.4°C). Q_{total}, Total flow rate. (**D**) Phantom-nerve cooling rates greater than 3°C/s are prompted by initiation of a 300 µl/min flow of PFP. (**E**) Systematic reduction in PFP flow rate from 300 µl/min enables controlled rewarming of the nerve over a 20 min period. (**F**) Evaporative microfluidic cooling enables precise and stable nerve cooling for more than 15 min. (**G**) The effect of a convective environment on the ability to cool a phantom nerve, as shown by a comparison of 37°C water and hydrogel baths. (**H**) Physiologically relevant rat sciatic nerve blood flow rates induce an increase in nerve temperature during cooling from 3° to 5°C in a phantom nerve. Error bars represent standard deviation of three trials. (**I**) POC microfluidics provide precise nerve cooling for 21 days in vitro (PBS, 37°C, pH 7.4).

and consistent nerve cooling to 3.0°C for a 15-min interval in Fig. 3F [minimum temperature (T_{min}) = 2.1°C; maximum temperature (T_{max}) = 3.5°C].

Thermal transfer in nerves occurs mainly through conduction, though subcutaneous biofluid and nerve blood flow contribute to convective thermal transfer. Comparisons of measurements in a conduction-only environment (hydrogel, 37°C) to those that include convective effects (water, 37°C) highlight these effects (Fig. 3G). The perfusion of blood through the targeted nerve presents an additional source of heat flux. For a nerve blood flow of 50 µl/min (Fig. 3H), the temperature of the nerve increases by 2.0°C (from 3.5° to 5.5°C). Experimental details are provided in fig. S8. A thermofluidic model (fig. S9) predicts the resultant changes in nerve temperature with and without nerve blood flow and shows similar results to experimental results (fig. S10). Details for the

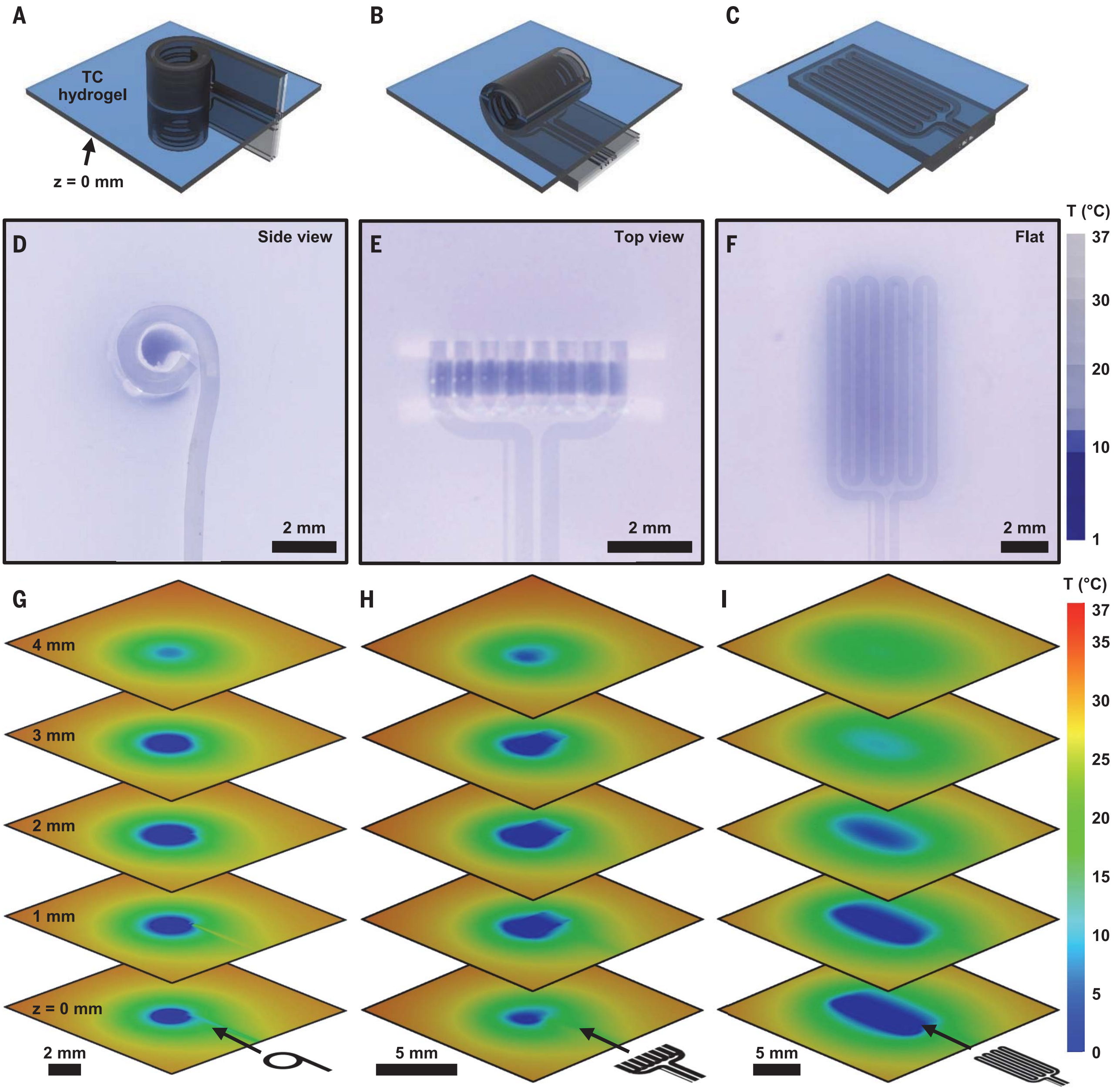


Fig. 4. Cooling localization. (**A** to **C**) Illustrations of the experimental setup for quantifying cooling localization (A) radially, (B) longitudinally, and (C) across the surface of the nerve. (**D** to **F**) A thermochromic (TC) hydrogel visually indicates the 10°C thermocline for a plane (D) perpendicular to the nerve, (E) parallel to the nerve, and (F) above a flat, uncurled cooler. (**G** to **I**) Simulations of the cases shown in (A) to (C), respectively, starting at the bottom of the TC hydrogel (z = 0).

thermofluidic model are provided in the supplementary materials. Data captured for daily cooling for 180 s over 21 consecutive days indicate stable operation in 37°C PBS, with a standard deviation in cooling temperature of 0.6°C (Fig. 3I).

Localization of the cooling effect to a predefined site without the need for insulating layers represents a key capability of the evaporative microfluidic cooling approach introduced here. Nerve coolers embedded in a thermochromic tissue mimic in three different configurations serve as models for experimental study of temperature gradients in radial (Fig. 4A) and longitudinal (Fig. 4B) views along the nerve and across the surface of an uncurled, planar device (Fig. 4C). Figure S11 illustrates the experimental setup for quantifying the 10°C thermocline in similar views of curled devices and a top-down view of a flat device. The 10°C thermocline is contained within the cuff (Fig. 4D) except for a region that extends downward ~500 μm from the exterior of the cuff and does not extend past the edge of the device longitudinally along the nerve (Fig. 4E). Figure 4F demonstrates that this thermocline extends over nearly the entire serpentine evaporation chamber in the flat device. Images of flow rate sweep experiments for the cases shown in Fig. 4, E and F, are in figs. S12 and S13, respectively. Thermal three-dimensional finite element analysis confirms that the cooling effect is largely confined radially (Fig. 4G) and longitudinally (Fig. 4H) within the extent of the cooling cuff and above a flat cooler (Fig. 4I). The temperature of the vapor remains confined inside the microfluidic channel, as indicated by the cold region that extends radially down and to the right in the z = 0 mm plane for Fig. 4G.

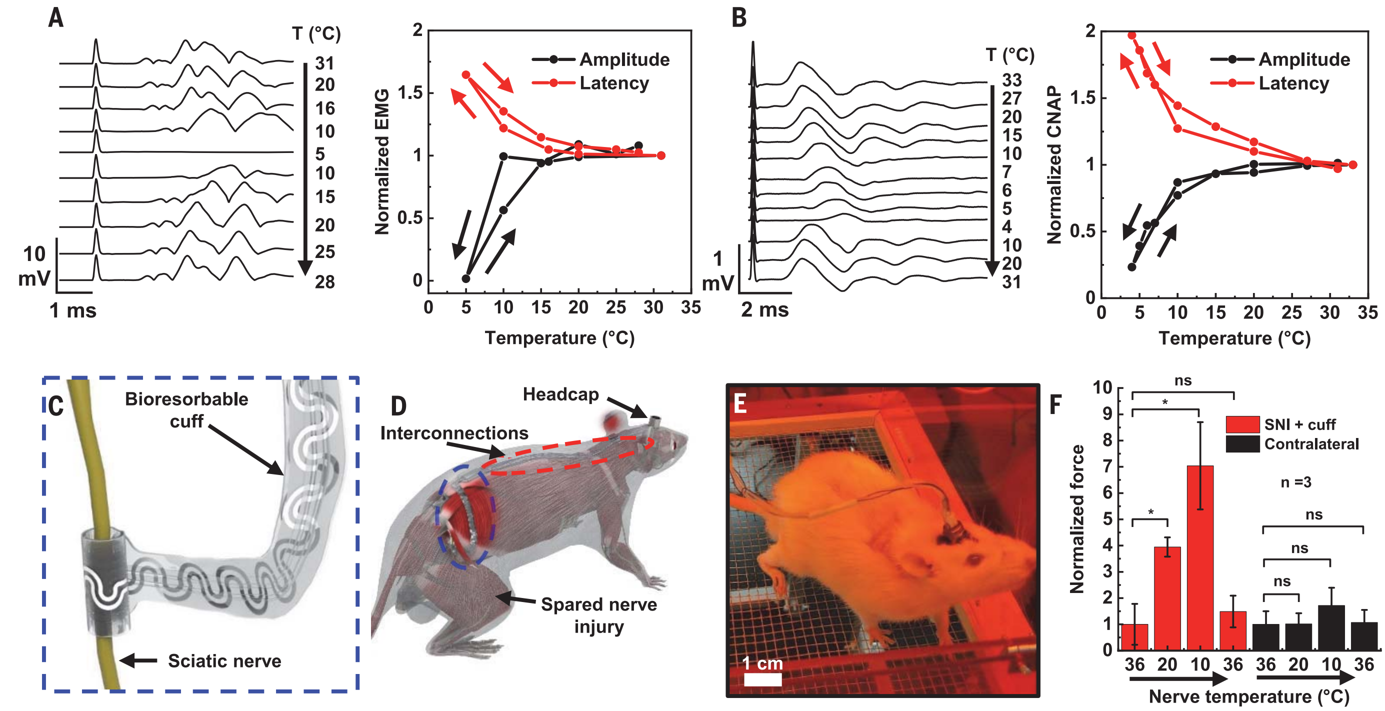


Fig. 5. Cooling-induced nerve block and analgesia. (**A**) Amplitude of muscular contraction in the tibialis anterior, as measured by EMG, diminishes with reduced temperature, and signal latency increases and subsequently recovers after rewarming. (**B**) Amplitude of CNAP diminishes with reduced temperature, and signal latency increases and subsequently recovers upon rewarming. (**C**) Illustration of the bioresorbable nerve cooler interfacing with the sciatic nerve of a rat. (**D**) Illustration of the subcutaneous routing path of the bioresorbable device and nonbioresorbable interconnects. A SNI generates persistent touch sensitivity of the paw and serves as a model for neuropathic pain. (**E**) Image of a freely moving rat in a test environment for quantifying the mechanical nociceptive threshold. (**F**) Cooling the sciatic nerve to 10°C increases the mechanical nociceptive threshold by a factor of seven in animals that received both a SNI and cuff, indicating a significant cooling-induced analgesic effect. Changes in mechanical sensitivity of the contralateral for the same animals are insignificant. Error bars represent standard error of the mean. ns, not significant; $^*P < 0.05$.

The low heat capacity of the PFP and N_2 exhaust gas mixture yields minimal cooling along the exterior of the outlet channel, as supported by experiments and simulations (figs. S14 and S15).

Acute animal trials demonstrate the capability of evaporative microfluidic coolers to reversibly eliminate evoked nerve signals. The soft, curled structure defines an intimate thermal and mechanical interface to the rat sciatic nerve without sutures and without mechanically induced damage (fig. S16). Electromyography (EMG) of the tibialis anterior muscle indicates a 92% reduction in EMG magnitude and 64% increase in signal latency of neuromuscular activity during cooling from 31° to 5°C over a period of 8 min (Fig. 5A). The amplitude and latency return to 108 and 100% of their initial values, respectively, after rewarming over a period of 3 min. Electrical recording from the sciatic nerve distal to the cooling cuff provides a measure of compound nerve action potential (CNAP) evoked through a single stimulation pulse. Cooling from 33° to 4°C over a period of 15 min prompts a decrease in signal amplitude of 77% and increase in latency of 97% (Fig. 5B). Amplitude and latency return to within 101 and 97% of their initial values, respectively, after subsequent rewarming over a period of 3 min. Examples of in vivo cooling profiles are provided in fig. S17.

Experiments based on cooling of sciatic nerves in a rat model of neuropathic pain associated with spared nerve injury (SNI) elucidate capabilities for microfluidic evaporative coolers to block pain in freely moving animals (*29*). In devices constructed for multiday experiments, transcutaneous connections to a bioresorbable microfluidic evaporative cooler and temperature sensor mounted to the sciatic nerve (Fig. 5C) route subcutaneously along the spine to a headcap (Fig. 5D). Device details are provided in figs. S18 and S19. An integrated connecter mounted inside a titanium headcap enables reversible fluidic and electronic connection to an awake, freely moving animal (Fig. 5E). Mechanical nociceptive sensitivity tests in two control animals (SNI only) over 3 weeks after the SNI show an expected increase in the mechanical sensitivity threshold of the SNI side as compared with that of the contralateral side, which persists for more than 3 weeks (fig. S20). Studies using three additional animals that received both the SNI and the cooling cuff show that cooling of the SNI-treated nerve from 37° to 10°C after 3 weeks of implantation leads to a sevenfold increase in the mechanical sensitivity threshold (1.6 to 11.5 g), consistent with a significant cooling-induced analgesic effect (Fig. 5F). Changes in the mechanical sensitivity threshold in the contralateral side during cooling of the SNI-treated nerve are statistically insignificant. Histologic analyses after 1, 2, 3, and 6 months indicate a close proximity of the cuff to the nerve and provide evidence of biocompatibility and bioresorption, as described in the supplementary materials (figs. S21 to S25 and table S1). Previous in vivo studies of POC (*21*), Mg (*30*), SiO_2 (*31*, *32*), and cellulose acetate (*33*) provide additional strong evidence of biocompatibility and associated bioresorption processes.

The results presented here demonstrate that spatiotemporally precise cooling enabled by soft, bioresorbable evaporative microfluidics provides for reversible elimination of local peripheral nerve activity. Experiments indicate capabilities in on-demand analgesia for the management of neuropathic pain in freely moving animal models. These concepts, materials, and device designs establish the engineering foundations for a class of implantable cooling systems capable of targeted neural blocking with relevance across a range of

clinical applications, including targeted, on-demand, nonopioid pain management.

REFERENCES AND NOTES

1. J. M. Hah, B. T. Bateman, J. Ratliff, C. Curtin, E. Sun, *Anesth. Analg.* **125**, 1733–1740 (2017).
2. K. Yu, X. Niu, B. He, *Adv. Funct. Mater.* **30**, 1908999 (2020).
3. T. R. Deer *et al*., *Bioelectron. Med.* **6**, 9 (2020).
4. R. Melzack, P. D. Wall, *Science* **150**, 971–979 (1965).
5. M. J. Cousins, D. B. Carr, T. T. Horlocker, P. O. Bridenbaugh, Eds., *Cousins & Bridenbaugh's Neural Blockade in Clinical Anesthesia and Pain Medicine* (Lippincott Williams & Wilkins, 2009).
6. S. M. Iyer *et al*., *Nat. Biotechnol.* **32**, 274–278 (2014).
7. C. Rabut *et al*., *Neuron* **108**, 93–110 (2020).
8. V. B. Brooks, in *Reviews of Physiology, Biochemistry and Pharmacology, Volume 95*, R. H. Adrian *et al*., Eds. (Springer, 1983), pp. 1–109.
9. E. H. Lothet *et al*., *Sci. Rep.* **7**, 3275 (2017).
10. R. Janssen, *Neurosci. Biobehav. Rev.* **16**, 399–413 (1992).
11. S. B. Rutkove, *Muscle Nerve* **24**, 867–882 (2001).
12. P. Borgdorff, P. G. A. Versteeg, *Eur. J. Appl. Physiol. Occup. Physiol.* **53**, 175–179 (1984).
13. Z. Zhang *et al*., *J. Neurophysiol.* **115**, 1436–1445 (2016).
14. T. Morgan *et al*., *J. Neurophysiol.* **123**, 2173–2179 (2020).
15. D. M. Ackermann, E. L. Foldes, N. Bhadra, K. L. Kilgore, *J. Neurosci. Methods* **193**, 72–76 (2010).
16. J. Jia, M. Pollock, *Muscle Nerve* **22**, 1644–1652 (1999).
17. J. Jia, M. Pollock, J. Jia, *Brain* **121**, 989–1001 (1998).
18. J. Shin *et al*., *Nat. Biomed. Eng.* **3**, 37–46 (2019).
19. J. Koo *et al*., *Nat. Med.* **24**, 1830–1836 (2018).
20. G. H. Borschel, K. F. Kia, W. M. Kuzon Jr., R. G. Dennis, *J. Surg. Res.* **114**, 133–139 (2003).
21. J. Yang, A. R. Webb, G. A. Ameer, *Adv. Mater.* **16**, 511–516 (2004).
22. R. T. Tran, J. Yang, G. A. Ameer, *Annu. Rev. Mater. Res.* **45**, 277–310 (2015).
23. R. T. Tran *et al*., *J. Biomed. Mater. Res. A* **102**, 2793–2804 (2014).
24. P. G. A. Rogueda, *Drug Dev. Ind. Pharm.* **29**, 39–49 (2003).
25. P. Grayburn, *Clin. Cardiol.* **20**, I12–I18 (1997).
26. M. Castrén *et al*., *Circulation* **122**, 729–736 (2010).
27. O. D. Kripfgans, J. B. Fowlkes, D. L. Miller, O. P. Eldevik, P. L. Carson, *Ultrasound Med. Biol.* **26**, 1177–1189 (2000).
28. P. A. Mountford, M. A. Borden, *Adv. Colloid Interface Sci.* **237**, 15–27 (2016).
29. I. Decosterd, C. J. Woolf, *Pain* **87**, 149–158 (2000).
30. X. Gu, Y. Zheng, Y. Cheng, S. Zhong, T. Xi, *Biomaterials* **30**, 484–498 (2009).
31. S.-W. Hwang *et al*., *Science* **337**, 1640–1644 (2012).
32. S.-K. Kang *et al*., *Nature* **530**, 71–76 (2016).
33. M. N. Nosar *et al*., *Cellulose* **23**, 3239–3248 (2016).

ACKNOWLEDGMENTS

Funding: This work was funded by the Phil and Penny Knight Campus for Accelerating Scientific Impact (J.T.R.); the Querrey Simpson Institute for Bioelectronics (J.A.R.); National Science Foundation grant CMMI1635443 (Y.H.); National Natural Science Foundation of China grant 12072057 (Z.X.); LiaoNing Revitalization Talents Program grant XLYC2007196 (Z.X.); Dalian Outstanding Young Talents in Science and Technology grant 2021RJ06 (Z.X.) Fundamental Research Funds for the Central Universities grant DUT20RC(3)032 (Z.X.); and National Research Foundation of Korea grant NRF-2021R1A5A1032937 (M.-H.S.). This work used the Northwestern University Micro/Nano Fabrication Facility (NUFAB), which is partially supported by the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), the Materials Research Science and Engineering Center (DMR-1720139), the State of Illinois, and Northwestern University. Histology services were provided by the Northwestern University Mouse Histology and Phenotyping Laboratory, which is supported by National Cancer Institute grant P30-CA060553 awarded to the Robert H. Lurie Comprehensive Cancer Center. **Author contributions:** Conceptualization: J.T.R., J.A.R.; Data curation: J.T.R., Z.X.; Funding acquisition: W.Z.R., Y.H., M.R.M., J.A.R.; Investigation: J.T.R., Z.X., Q.Y., M.-H.S., Y.Y., Y.D., K.R.J., C.L., S.M., E.P., A.J., Z.Z., M.J.K., Y.X., I.H., R.A., C.F., X.G.; Methodology: J.T.R., Z.X., Q.Y., M.-H.S., Y.Y., S.R.K., Y.D., Y.H., C.L., C.F., W.Z.R., Y.H., M.R.M., J.A.R.; Project administration: J.T.R., Z.X., Y.Y., K.R.J.; Visualization: J.T.R.; Writing – original draft: J.T.R., Z.X., E.R., J.A.R.; Writing – review & editing: J.T.R., Z.X., Q.Y., S.R.K., J.A.R. **Competing interests:** One or more provisional patents are being filed on this work. **Data and materials availability:** All data are available in the main text or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl8532
Materials and Methods
Supplementary Text
Figs. S1 to S29
Table S1
References (*34–52*)

Submitted 9 August 2021; accepted 13 May 2022
10.1126/science.abl8532

PALEOCLIMATE

Tectonic degassing drove global temperature trends since 20 Ma

Timothy D. Herbert[1]*†, Colleen A. Dalton[1]†, Zhonghui Liu[2], Andrea Salazar[3], Weimin Si[1], Douglas S. Wilson[4]

The Miocene Climatic Optimum (MCO) from ~17 to 14 million years ago (Ma) represents an enigmatic reversal in Cenozoic cooling. A synthesis of marine paleotemperature records shows that the MCO was a local maximum in global sea surface temperature superimposed on a period from at least 19 Ma to 10 Ma, during which global temperatures were on the order of 10°C warmer than at present. Our high-resolution global reconstruction of ocean crustal production, a proxy for tectonic degassing of carbon, suggests that crustal production rates were ~35% higher than modern rates until ~14 Ma, when production began to decline steeply along with global temperatures. The magnitude and timing of the inferred changes in tectonic degassing can account for the majority of long-term ice sheet and global temperature evolution since 20 Ma.

The climatic journey from global warmth to the great Pleistocene ice ages began many tens of millions of years ago, in the late Eocene (~40 Ma) when global temperatures began to descend from the strikingly warm conditions of the Mesozoic and Early Cenozoic (*1*). This path was far from monotonic, however. The climate system reached a tipping point at the Eocene/Oligocene boundary, when a substantial icecap grew on Antarctica and global temperatures declined. Episodic advances and retreats of this Southern Hemisphere icecap occurred for most of the Oligocene and Miocene (*2*, *3*). However, an important reversal of the climatic trend occurred during the Miocene Climatic Optimum (MCO), when the East Antarctic Ice Sheet (EAIS) may have largely disappeared (*4*, *5*).

The MCO was originally recognized by a substantial excursion to lighter values in the benthic oxygen isotope record, which could reflect a rise in deep sea temperatures or a decrease in global ice volume (*1*, *2*, *6*, *7*). Although the partitioning of the two components remains challenging, it is clear that both warming and deglaciation occurred on and around the Antarctic continent (*4*, *5*, *8*). In Fig. 1, we provide a continuous global estimate of marine temperatures through the MCO by moving to higher latitudes where the pattern of temperature change is clear, although temperature changes are certainly amplified relative to a global mean ocean temperature. The curves rely almost entirely on the alkenone proxy, with a single exception in the Southern Hemisphere between an alkenone-based record at Deep Sea Drilling Project Site 594 (*9*) spliced to a continuous GDGT (glycerol dialkyl glycerol tetraether)–based estimate at Integrated Ocean Drilling Program Site 1352 over the MCO (*10*).

This synthesis of Miocene ocean temperatures indicates that the MCO was a local thermal maximum superimposed on a strikingly warm background climate. In comparison to the present day, temperatures preceding and postdating the MCO were high for millions of years: A global alkenone-based reconstruction at 10 Ma estimated an area-weighted ocean warming of +6°C relative to the present day (*9*); Fig. 1 suggests that temperatures prior to the MCO may have been similar. Estimating the global temperature anomaly during the MCO itself is difficult, because the temperatures were so warm that the alkenone proxy became saturated over a band of ± ~40° latitude around the equator, implying temperatures in excess of ~29°C. However, by relying on regional Northern and Southern Hemisphere stacks where paleotemperatures remain below alkenone saturation, we arrive at an estimate of a global mean surface air temperature (SAT) anomaly of ~ +12° to 19°C using a scaling of ~1.7 times (*11*) the global sea surface temperature (SST) anomaly of +7.25° to 11.5°C (fig. S1). The warmth we estimate is on the high side of a recent estimate of MCO ΔSAT (+11.5°C) compiled from heterogeneous SST proxies, which includes tropical estimates of perhaps questionable reliability (*12*).

That the MCO represents a local maximum in global temperatures is also confirmed by high-resolution Mg/Ca records showing clear warming into the MCO that parallels the marine $\delta^{18}O$ curve at two tropical locations (*13*) and pronounced cooling at the end of the MCO at South Pacific ODP Site 1171 (*14*), supported by similar cooling in a GDGT-based record from the Arabian Sea (*15*). The link of inferred warming to deglaciation in Antarctica is supported by micropaleontological and sedimentological observations of open water surrounding much of the Antarctic during the MCO (*5*), with vegetation occupying at least coastal areas of that continent during this time (*8*). A warm high-latitude Southern Ocean is also consistent with clumped isotope results on bottom-dwelling foraminifera that indicate MCO deep ocean temperatures ~10°C warmer than at present (*16*).

It is thus evident that Miocene warmth was global, long-lived, and peaked with the retreat of ice over much of Antarctica. Global cooling occurred with the reoccupation of much of the Antarctic by ice after the MCO, accelerating after ~ 9.5 Ma (Fig. 1).

Here, we investigated the possibility that global climate since 20 Ma was paced and synchronized by changes in a slow, but powerful, deep Earth process: the rate of degassing of CO_2 controlled by changes in the rate of oceanic plate creation and destruction. Our hypothesis builds on seminal work by Berner *et al.* (*17*), who proposed that on time scales longer than the residence time of carbon in the ocean-atmosphere-biosphere system (>100,000 years),

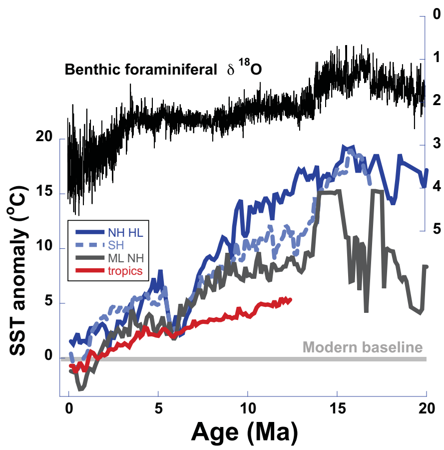


Fig. 1. Regionally averaged marine sea surface temperature anomalies relative to the present day, compared to the evolution of oxygen isotopic values of bottom-dwelling foraminifera. Data are from (*7*). Abbreviations: NH HL, Northern Hemisphere high latitude (45° to 69°N); SH, Southern Hemisphere (20° to 49°S); ML NH, mid-latitudes of the Northern Hemisphere (29° to 43°N); Tropics, all tropical sites (24°N to 2°S). All paleotemperature determinations were made with the alkenone proxy, with the exception of Site 1352 (part of the Southern Hemisphere stack), where we rely on GDGT-based paleothermometry (data S1 and S2). Marine temperatures evolve essentially synchronously in both hemispheres and reach a local maximum at the time of the MCO. The large cooling from ~9.5 to 6 Ma is not reflected in the benthic record, presumably because there was little growth of continental ice during this period.

[1]DEEPS, Brown University, Providence, RI 02912, USA. [2]Department of Earth Sciences, University of Hong Kong, Hong Kong, China. [3]Department of Earth and Planetary Science, Harvard University, Cambridge, MA, USA. [4]Department of Earth Science, University of California, Santa Barbara, CA, USA.
*Corresponding author. Email: timothy_herbert@brown.edu
†These authors contributed equally to this work.

the inventory of CO_2 in the atmosphere would be controlled by a plate tectonic degassing source term, buffered by the negative feedback of silicate weathering (*18*, *19*). Most subsequent treatments of the long-term carbon cycle have instead focused on removal rates of carbon as the forcing function, often explicitly on the assumption that tectonic degassing has not changed appreciably (*19*–*23*). However, investigating the tectonic forcing of climate in the past 20 million years offers several key advantages over studies focused on longer time periods: The motions of tectonic plates are well known because of the preservation of young seafloor, magnetic polarity reversals are dated with high accuracy and small uncertainties, and large and well-dated transitions occurred in global climate.

We derived a global synthesis of ocean crust production rate, which we propose is a proxy for degassing rates over time (*24*). This assumption does not specify the proportion of degassing that occurs at mid-ocean ridges (*25*), in metamorphism in orogenic belts (*26*, *27*), or by release in subduction zones (*28*–*30*), as faster crust production necessarily implies faster lithosphere destruction (and the converse). Our simple model neglects factors that are likely to be important in detail, such as the carbon content of sediments and rocks fed into orogenic and subduction zones, the dip of the subducting plate, and other factors that might influence the efficiency of tectonic degassing independently of the total rate of marine crustal recycling (*28*, *31*).

Crustal production rates are calculated as the product of spreading rate and ridge length, assuming that crustal thickness is constant in space and time. Compilations of ocean crustal thickness have shown that it is largely independent of spreading rate except at ultraslow spreading ridges where melt supply is low (*32*). We took advantage of recent reconstructions of ocean ridge spreading histories at high temporal resolution (*33*–*35*) as well as new determinations of spreading rates in the eastern Pacific. Central to our analysis is a careful treatment of three distinct sources of error: in the rotation parameters that describe seafloor spreading, in the endpoints of each mid-ocean ridge, and in the ages of magnetic polarity reversals.

The globally integrated crustal production curve reveals large changes since ~20 Ma (*24*) (Fig. 2A). Total crustal production decreased from at least 3.3 km^2/year prior to 14 Ma to no more than 2.7 km^2/year after 6 Ma. Ridges in the eastern Pacific dominate total production (Fig. 2, B and C), but crust production has slowed along most ridges since 20 Ma. Only the Pacific-Antarctic ridge substantially increased production; ridges in the eastern Pacific reduced production by 25 to 50%, and those in the Atlantic slowed by 10 to 30%.

Previous studies of ocean crust production differ in their estimate of the trend since the mid-Miocene, ranging from increasing (*36*) to constant (*37*) to decreasing (*38*, *39*). Among the subset of studies that identified declining ridge flux, our result has a larger amplitude, greater level of detail, and better-defined uncertainty. Three main factors may explain these improvements: an updated reconstruction of the complex and incomplete record of spreading in the eastern Pacific since 25 Ma, the inclusion of ridge spreading histories with higher temporal resolution, and our use of astronomically dated magnetic polarity ages (*40*, *41*). Important pieces of the record of relative motion of the Cocos, Nazca, and Pacific plates are obscured by large microplates, poor low-latitude magnetic geometry, and hotspot tracks. Moreover, some studies (*36*, *37*, *42*) derive spreading rates from gridded estimates of seafloor age grid, early versions of which have been shown to contain some errors in the eastern Pacific for ages of 5 to 25 Ma (*43*). Although we compiled the eastern Pacific spreading record in more detail than previous authors, we caution that subdividing the interval from 22 to 15 Ma remains difficult because of the limited record of the Nazca-Pacific and Cocos-Nazca plate pairs. Other studies derive spreading rates from rotation poles, as we do, but are focused on a longer time period and typically consider only two to four intervals within the past 20 Ma (*36*, *39*), in contrast to our use of 10 intervals (table S3). Furthermore, relative to the CK95 time scale (*44*) used by many earlier studies, the astronomical polarity time scale used here (*40*) importantly shortens the interval from C5C to C6 (~16 to 19 Ma). The smaller errors in the astronomical dating of magnetic ages also reduce uncertainties in crust production rate and allow greater confidence to be assigned to the rate changes we measured (data S3).

The similarity of the new crust production rates to the broad evolution of ocean temperatures motivated us to evaluate the plausibility of two end-member explanations for cooling since the Miocene, one focused on changes in continental weatherability (CO_2 sink) and the other on CO_2 degassing (CO_2 source). The weatherability argument posits a long-term increase in the proportion of chemical weathering in relation to continental breakdown (*19*, *22*) or to the emergence of maritime continents with easily weathered lithologies in the tropical Indo-Pacific in the Plio-Pleistocene (*23*).

To explore the alternative (source) scenario, we constructed a carbon cycle mass balance forced by the new crust production spreading curve, which is used as a proxy for tectonic carbon release (supplementary materials). Because nearly all plate pairs display a pattern similar to that of the global ensemble (Fig. 2A), we made the simplification that tectonic degassing scales linearly to the global ocean crustal production function. Two adjustable parameters can be optimized to satisfy the twin constraints of CO_2 degassing rates and global temperatures: (i) the Earth system sensitivity that relates atmospheric CO_2 to long-term equilibrium global temperature (ESS), and (ii) the sensitivity of the silicate weathering negative feedback to temperature (*Ea*), which is believed to keep the carbon system in long-term balance. There is good reason to suspect that ESS is larger than a recent estimate (*45*) of 2.6° to 3.9°C per doubling of CO_2 [referred to as equilibrium climate

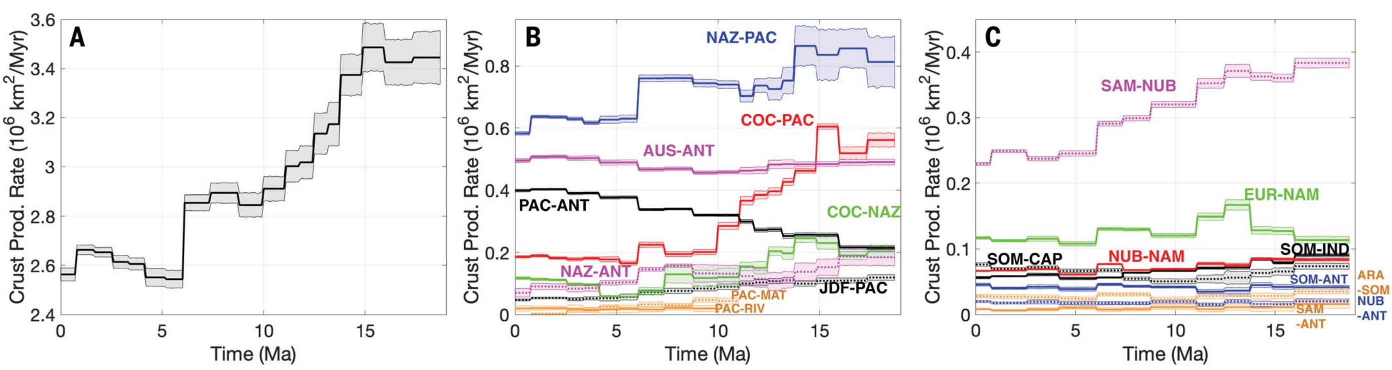


Fig. 2. Ocean crust production rates with 95% confidence interval. See data S3. (**A**) Global total. (**B** and **C**) By plate boundary. Note the different vertical scales for (B) and (C). See table S2 for plate names.

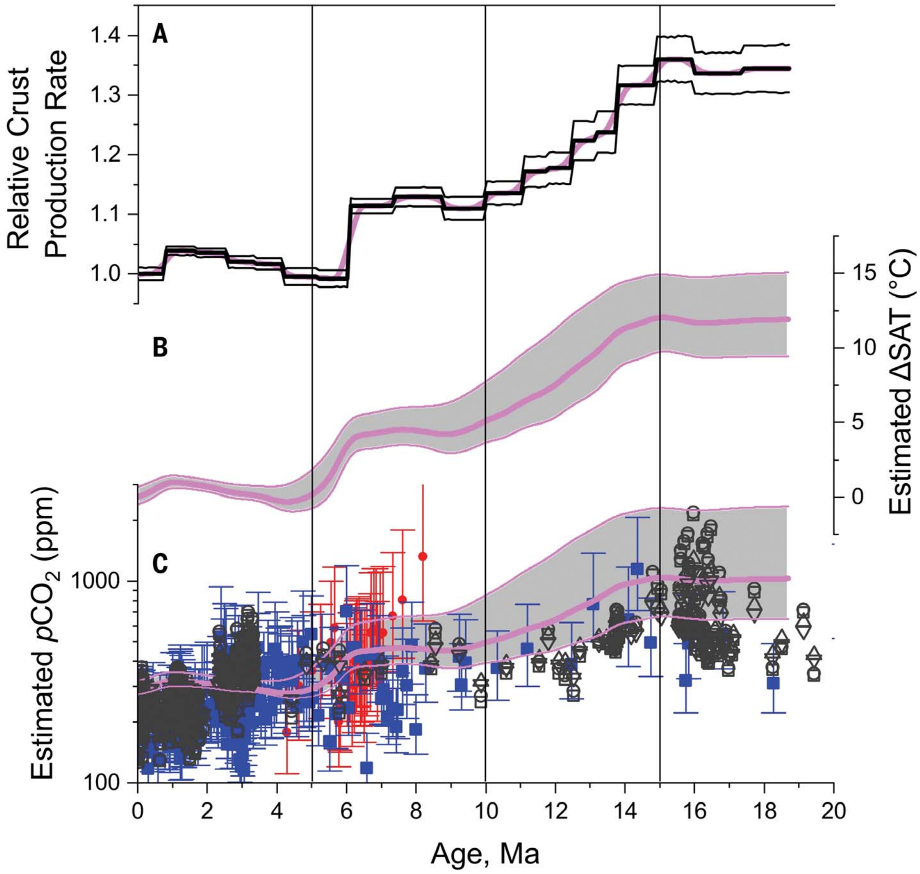


Fig. 3. Results of carbon mass balance modeling. (**A**) Global degassing rate driven by ocean crustal production reconstruction. (**B**) Calculated changes in global surface air temperature (SAT). (**C**) Corresponding implied atmospheric CO_2 levels. Thick pink lines in (B) and (C) result from using central crustal production estimate [thick line in (A)]. Central carbon cycle parameters are *Ea* = 4 kcal/mol and ESS = 6.5°C per doubling of CO_2. Upper and lower bounds in (B) and (C) vary Ea from 3.5 to 4.5 kcal/mol and ESS from 5° to 8°C. Gray/black symbols in (C) show pCO_2 estimates from boron isotope analyses of planktonic foraminifera using different ocean carbon system parameters (*49*); blue and red symbols are based on carbon isotope values of alkenones (*48*, *50*).

sensitivity (ECS or "Charney" climate sensitivity)] and also to consider the possibility that ESS may increase nonlinearly under high CO_2 (*11*, *46*). We explored the effects of varying ESS between 5° and 8°C per CO_2 doubling, or approximately two times the 66th percentile bounds of the ECS (*45*). On tectonic time scales, ocean total dissolved inorganic carbon and alkalinity are primarily set by degassing, chemical weathering, and calcium concentration of seawater; therefore, it is not necessary to know the absolute values of these quantities a priori to estimate equilibrium pCO_2 (supplementary materials).

The temperature dependence of weathering sensitivity is directly constrained by our estimate of peak MCO warmth and degassing. Here, *Ea* represents an effective weathering dependence on temperature in an Arrhenius formulation; it is a simple parameterization of the complex factors that control silicate weathering, including relief, the insulating effect of soil formation (transport limitation), and rainfall. The values chosen here, although considerably lower than those derived from pure silicate mineral activation energies, are consistent with previous studies of global weathering rates (supplementary materials).

The model satisfies the observation of generally warm early to mid-Miocene conditions, the timing of the MCO, and the steep cooling observed globally in the younger half of the Miocene (14.5 to 6 Ma). The level of detail in the crustal production curve from ~14 to 20 Ma is limited by time resolution in tectonic constraints (rotation points at 13.739, 15.974, and 18.748 Ma) and by somewhat higher uncertainties in the astronomical calibration of the magnetic polarity time scale. Degassing rates may have changed very little since ~6 Ma (Fig. 2). If our proposed linkage between seafloor spreading and tectonic degassing is correct, it carries several major implications. First, the similarity in timing between crustal production changes and paleotemperatures suggests a geologically efficient (million-year time scale) coupling between tectonic processes and CO_2 release to the Earth surface; time lags between tectonic forcing and climate response are not apparent at the resolution of our analysis. Second, it suggests that many proxy-based reconstructions of Miocene levels of CO_2 may be too low (Fig. 3C). The large range in plausible CO_2 levels in our reconstruction derive from the wide permissible range of ESS. However, it is very difficult to reconcile the combination of peak Miocene warmth and high degassing with an atmospheric CO_2 inventory much lower than 850 ppm (*12*) (see Fig. 3). In this regard, we note that the trend of more recent CO_2 proxy estimates has been to revise reconstructed Miocene levels of CO_2 higher than earlier estimates (*47*–*50*), although a gap remains between our preferred CO_2 curve (Fig. 3C, heavy pink line) and some proxy estimates. Last, our finding of large variations in seafloor spreading rate over the Neogene implies that interpretations of geochemical signals commonly used in paleoceanography (Mg/Ca, $^{87}Sr/^{87}Sr$) may need to be reinterpreted in light of changing hydrothermal fluxes driven by changes in seafloor spreading rates (*19*, *51*).

Forcing by a short-lived and relatively small input of CO_2 from Columbia River basalt magmatism (*52*) appears insufficient as a first-order explanation for the long duration of a strikingly warm Miocene climate, although it may have contributed to the apex during the MCO. In contrast, the hypothesis of long-term, large variations in tectonic degassing provides a potent, persistent driver. The evolution of temperature, inferred CO_2 levels, and benthic $\delta^{18}O$ values (Figs. 1 and 3) suggests that polar glaciation exhibits strongly nonlinear threshold sensitivity to CO_2 and temperature: The large retreat of the East Antarctic Ice Sheet during the MCO may have required only a modest rise in CO_2 from its generally high early Miocene level, and the onset of Northern Hemisphere ice ages at ~2.7 Ma occurred under tectonic CO_2 forcing similar to today. These results are consistent with two very different CO_2/temperature thresholds for glaciation in the Southern and Northern hemispheres

proposed by (*53*) and especially imply a high sensitivity of East Antarctic ice volume to global temperature at a "set point" of ~ +10°C (Fig. 3).

We do not rule out the possibility that the global carbon cycle was also affected by a transition from weaker silicate "weatherability" in the Miocene relative to the Plio-Pleistocene (*22*, *23*). However, the proposed low weatherability cannot serve as the primary explanation for the warm early to mid-Miocene: In the case of the scenario evaluated by Park *et al.* (*23*), a reduced MCO weatherability accounts for about one-fourth of the observed temperature anomaly (fig. S3). A 30% higher tectonic CO_2 flux rate will, however, generate enough warming to be consistent with our empirical SST and SAT estimates (Fig. 3). We readily admit that our temperature and CO_2 model reconstructions are imperfect, because our assumptions that CO_2 degassing rate is in 1:1 proportion to ocean crustal production, and that weatherability and ESS remain fixed for the past 20 million years, are unlikely to apply strictly. In this regard, reduced weatherability during the mid-Miocene (*22*) may have enhanced large excursions in atmospheric CO_2 and temperature by reducing the sensitivity of the silicate weathering sink in a warm climate in comparison to the present day.

Our findings suggest that the coupled evolution of CO_2 and climate over the Neogene can be understood as a combination of exogenous, slow forcing by the tectonic CO_2 source, in combination with more rapid but substantial repartitioning of the total surface carbon inventory among atmosphere, biosphere, and ocean paced by orbital cycles. Our study does not negate the rich possibility of other tectonic forcings of the carbon cycle (*19*–*23*, *29*, *54*), but it suggests that the dynamism in ocean crustal creation/destruction as a carbon cycle driver has been underestimated since the publication of (*17*). Riding on top of tectonic forcing, cyclical sequestration and release of carbon from the deep ocean (*55*) may explain high-frequency orbital-scale variability evident within the MCO, as demonstrated by cyclic variations in foraminiferal $\delta^{18}O$-, $\delta^{13}C$-, and Mg/Ca-based SST estimates (*4*, *6*, *13*, *14*, *56*); processes driving the oceanic carbon reservoir became dominant during the Pleistocene ice ages.

REFERENCES AND NOTES

1. J. Zachos, M. Pagani, L. Sloan, E. Thomas, K. Billups, *Science* **292**, 686–693 (2001).
2. B. P. Flower, J. P. Kennett, *Paleoceanography* **10**, 1095–1112 (1995).
3. D. Liebrand *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3867–3872 (2017).
4. R. Levy, D. Harwood, F. Florindo, SMS Science Team, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3453–3458 (2016).
5. F. Sangiorgi *et al.*, *Nat. Commun.* **9**, 317 (2018).
6. A. Holbourn, W. Kuhnt, K. G. D. Kochhann, N. Andersen, K. J. Sebastian Meier, *Geology* **43**, 123–126 (2015).
7. D. De Vleeschouwer, M. Vahlenkamp, M. Crucifix, H. Pälike, *Geology* **45**, 375–378 (2017).
8. S. Warny *et al.*, *Geology* **37**, 955–958 (2009).
9. T. D. Herbert *et al.*, *Nat. Geosci.* **9**, 843–847 (2016).
10. J. Jiang, thesis, University of Hong Kong (2019).
11. D. J. Lunt *et al.*, *Nat. Geosci.* **3**, 60–64 (2009).
12. N. J. Burls *et al.*, *Paleoceanogr. Paleoclimatol.* **36**, e2020PA004054 (2021).
13. S. M. Sosdian, C. H. Lear, *Paleoceanogr. Paleoclimatol.* **35**, e2020PA003920 (2020).
14. A. E. Shevenell, J. P. Kennett, D. W. Lea, *Science* **305**, 1766–1770 (2004).
15. G. Zhuang, M. Pagani, Y. G. Zhang, *Geology* **45**, 655–658 (2017).
16. S. E. Modestou, T. J. Leutert, A. Fernandez, C. H. Lear, A. N. Meckler, *Paleoceanogr. Paleoclimatol.* **35**, e2020PA003927 (2020).
17. R. A. Berner, A. C. Lasaga, R. M. Garrels, *Am. J. Sci.* **283**, 641–683 (1983).
18. J. C. Walker, P. Hays, J. F. Kasting, *J. Geophys. Res.* **86**, 9776–9782 (1981).
19. L. R. Kump, M. A. Arthur, in *Tectonic Uplift and Climate Change* (Springer, 1997), pp. 399–426.
20. M. Raymo, W. F. Ruddiman, *Nature* **359**, 117–122 (1992).
21. G. Li, H. Elderfield, *Geochim. Cosmochim. Acta* **103**, 11–25 (2013).
22. J. K. Caves Rugenstein, D. E. Ibarra, F. von Blanckenburg, *Nature* **571**, 99–102 (2019).
23. Y. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25319–25326 (2020).
24. C. A. Dalton, D. S. Wilson, T. D. Herbert, *Geophys. Res. Lett.* **49**, e2022GL097937 (2022).
25. A. E. Saal, E. H. Hauri, C. H. Langmuir, M. R. Perfit, *Nature* **419**, 451–455 (2002).
26. J. A. Becker, M. J. Bickle, A. Galy, T. J. Holland, *Earth Planet. Sci. Lett.* **265**, 616–629 (2008).
27. A. Skelton, *Geology* **39**, 43–46 (2011).
28. C.-T. A. Lee *et al.*, *Geosphere* **9**, 21–36 (2013).
29. N. R. McKenzie *et al.*, *Science* **352**, 444–447 (2016).
30. P. B. Kelemen, C. E. Manning, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3997–E4006 (2015).
31. T. Plank, C. E. Manning, *Nature* **574**, 343–352 (2019).
32. R. S. White, D. McKenzie, R. K. O'Nions, *J. Geophys. Res.* **97**, 19683–19715 (1992).
33. S. Merkouriev, C. DeMets, *Geophys. J. Int.* **173**, 1064–1083 (2008).
34. C. DeMets, S. Merkouriev, *Geophys. J. Int.* **207**, 741–773 (2016).
35. C. DeMets, S. Merkouriev, S. Jade, *Geophys. J. Int.* **220**, 1149–1171 (2019).
36. C. J. Rowan, D. B. Rowley, *Geophys. J. Int.* **208**, 1173–1183 (2017).
37. D. B. Rowley, *Geol. Soc. Am. Bull.* **114**, 927–933 (2002).
38. C. P. Conrad, C. Lithgow-Bertelloni, *Geology* **35**, 29–32 (2007).
39. R. D. Müller *et al.*, *Tectonics* **38**, 1884–1907 (2019).
40. F. J. Hilgen, L. J. Lourens, J. A. van Dam, in *The Geologic Time Scale 2012*, F. M. Gradstein, J. G. Ogg, M. Schmitz, G. Ogg, Eds. (Elsevier, 2012), pp. 923–978.
41. A. J. Drury *et al.*, *Earth Planet. Sci. Lett.* **475**, 254–266 (2017).
42. J.-P. Cogné, E. Humler, *Geochem. Geophys. Geosyst.* **7**, Q03011 (2006).
43. D. B. Rowley, *Lithosphere* **11**, 21–43 (2019).
44. S. C. Cande, D. V. Kent, *J. Geophys. Res.* **97**, 13917–13951 (1992).
45. S. Sherwood *et al.*, *Rev. Geophys.* **58**, e2019RG000678 (2020).
46. A. Farnsworth *et al.*, *Geophys. Res. Lett.* **46**, 9880–9889 (2019).
47. S. M. Sosdian *et al.*, *Earth Planet. Sci. Lett.* **498**, 362–376 (2018).
48. H. M. Stoll *et al.*, *Quat. Sci. Rev.* **208**, 1–20 (2019).
49. J. W. B. Rae *et al.*, *Annu. Rev. Earth Planet. Sci.* **49**, 609–641 (2021).
50. T. Tanner, I. Hernández-Almeida, A. J. Drury, J. Guitián, H. Stoll, *Paleoceanogr. Paleoclimatol.* **35**, e2020PA003925 (2020).
51. J. A. Higgins, D. P. Schrag, *Earth Planet. Sci. Lett.* **416**, 73–81 (2015).
52. J. Kasbohm, B. Schoene, *Sci. Adv.* **4**, eaat8223 (2018).
53. R. M. Deconto *et al.*, *Nature* **455**, 652–656 (2008).
54. S. Brune, S. E. Williams, R. D. Müller, *Nat. Geosci.* **10**, 941–946 (2017).
55. D. Archer, A. Winguth, D. Lea, N. Mahowald, *Rev. Geophys.* **38**, 159–189 (2000).
56. A. Holbourn, W. Kuhnt, M. Schulz, J.-A. Flores, N. Andersen, *Earth Planet. Sci. Lett.* **261**, 534–550 (2007).

ACKNOWLEDGMENTS

We thank A. Lee, C. Carchedi, and J. Smith for assistance with seafloor magnetic anomaly studies. This research used samples and/or data provided by the International Ocean Discovery Program (IODP) and its predecessors. IODP is sponsored by NSF and participating countries under management of Joint Oceanographic Institutions (JOI) Inc. **Funding:** Supported by NSF grants 1635127 (T.D.H., W.S., C.A.D.) and 1545859 (T.D.H.), the Doherty Fund for Oceanography at Brown University (T.D.H., W.S.), Hong Kong Research Grants Council grant 17305019 (Z.L.), and the Leadership Alliance (A.S.). **Author contributions:** Conceptualization, T.D.H., C.A.D.; methodology, T.D.H., C.A.D., Z.L., A.S., W.S., D.W.; writing–original draft, T.D.H., C.A.D., Z.L., A.S., W.S., D.S.W.; writing–reviewing and editing, T.D.H., C.A.D., W.S., D.S.W. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data supporting the regional SST trends in Fig. 1 include previously published (*9*, *57*–*66*) data available at the PANGAEA data repository (www.pangaea.de). Data tables providing new TEX_{86}-based estimates of SST and bulk stable isotopes at Site U1352, the age model used for Site U1352, and the regional SST stacks presented in Fig. 1 are also archived at the PANGAEA data repository.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4353
Supplementary Text
Figs. S1 and S2
Tables S1 and S2
Data S1 to S4
Matlab Codes for Carbon Cycle Calculations
References (*57*–*86*)

Submitted 16 July 2021; accepted 3 May 2022
10.1126/science.abl4353

WORKING LIFE

By **Kasper Bonnesen**

Stepping out of my comfort zone

"Do you want to go to Sunday brunch?" It was a question I'd never dared ask an office mate, but I was feeling desperate. I'd relocated from Denmark to the United States 2 months earlier for a research abroad experience during my Ph.D. I hadn't made any friends by that point so I decided to ask another Ph.D. student in my office whether they'd meet up outside work. He looked at me, smiled, and then uttered "Yes." It was a little thing, but it was a victory. I had finally conquered my shyness and dared to step out of my comfort zone.

My university encourages Ph.D. students to spend time abroad, so I planned from the beginning to work at a U.S. university for 6 months under a collaborator of my adviser's. I was excited to travel, improve my English, and expand my scientific network. But I was also afraid that my shy, introverted self would keep me from getting to know people, leading to a level of sadness and loneliness that would eventually force me to abandon the adventure and go back to Denmark.



ILLUSTRATION: ROBERT NEUBECKER

"Even little interactions at work helped me feel part of a community."

It had happened twice before. The first time was immediately after medical school when I took a job at a small hospital in Greenland. I had always wanted to go to the island, and the challenge of living in a small, isolated community and doing my best to help people appealed to me. But I had barely put my feet on Greenlandic soil before something felt wrong. Despite working with open and friendly colleagues, I never felt I belonged and was lonely and sad. Two weeks after my start date, I told the hospital manager I needed to go back to Denmark. The second time was during the pandemic. I moved to Copenhagen because I'd always wanted to live there and because my Ph.D. adviser was fine with me working remotely. That relocation only lasted 3 months—again because of loneliness. I missed my friends and family who lived near my university, which is less than an hour's drive from my hometown.

When my departure date neared for my research abroad experience in Atlanta, I vowed that this time would be different. I wanted to make more of an effort to get to know people and feel comfortable in the community, and I thought it would be easier because my outgoing younger sister happened to be doing her Ph.D. in the same city.

Still, the initial few months were difficult. Finding myself in a foreign country without my habitual life terrified me. I felt I had to learn everything from scratch—where do I get my groceries? How do I get to work? If my sister was busy, who could I spend time with? It didn't help that students were largely away from campus for the first 2 months—first because classes weren't in session and later because Atlanta was hit by a wave of COVID-19.

My sister helped me come up with a plan. We decided that, once the COVID-19 situation eased, I would go to campus every day. There, I would try to engage in conversations with my colleagues—for instance, by greeting people I passed in the hallway or by asking other students how their weekends went while waiting for the coffee to brew. Every weekend, I would also plan outings with at least one person, whether it be going out for brunch, visiting a museum, traveling to another city, or going on a hike.

The advice made a huge difference—I suddenly felt much less lonely. Even little interactions at work helped me feel part of a community. Gradually, I built up a small group of friends—from both within and outside the office—whom I could call on to do things with, including the Ph.D. student I had invited to brunch. By the end of my time in Atlanta, I felt so comfortable that I actually had mixed feelings about going home.

I returned to Denmark in May, feeling fulfilled by my time abroad. I was able to travel to new places, meet interesting people, and learn new professional skills. I also came to appreciate the benefits of stepping outside my comfort zone. I encourage every young scientist to do a research stay abroad—and to have the courage to open up to people while you're there. You never know what you'll learn. ■

Kasper Bonnesen is a Ph.D. student at Aarhus University. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.